الإحصاء المطبق في
العلوم الإجتماعية



الإستاذ محمد بداوي
الإستاذ محمد دوة

اليازوري

الإحصاء المطبق في العلوم الإجتماعية
الإستاذ محمد بداوي
الإستاذ محمد دوة
اليازوري

الاحصاء المطبق في العلوم الاجتماعية

تأليف :

الأستاذ : محمد بداوي (أستاذ بجامعة الأغواط)

الأستاذ : محمد دوة (أستاذ بجامعة الأغواط)

قال الله تعالى: "سبحانك لا علم لنا إلا ما علمتنا إنك أنت العليم الحكيم ".

الإهداء: إلى أهلنا في غزة...أهل العزة...

قائمة المحتويات	
الموضوع	**رقم الصفحة**

المقدمة

الحمد لله رب العالمين ، والصلاة والسلام على خاتم النبيئين وسيد المرسلين، نبينا محمد الهادي الامين الذي بعثه الله رحمة للعالمين، وعلى أله وأصحابه وأنصاره وأتباعه ومن أهتدى بهديه وعمل بسنته إلى يوم الدين.

و بعد:

يتضمن هذا الكتاب دروس موجهة الى طلبة سنة ثانية ليسانس ديمغرافيا (علم السكان) ، وشعر المؤلفان بالحاجة لمثل هذا الكتاب من خلال تدريسهما في قسم علم الاجتماع ، ومن خلال إشرافهما على عدد من المذكرات، ويمكن أن يكون هذا الكتاب بما يحويه من أساليب إحصائية متنوعة ، وبما يتضمنه من أمثلة تطبيقية عديدة، ذو فائدة لقطاع واسع من القراء المهتمين بالإحصاء.

إن هذا الكتاب كأي نتاج علمي لا يخلو من النواقص والهفوات، وكل أملنا أن يسهم في تطوير البحث العلمي.

ونسأل الله أن يجعل هذا العمل خالصا لوجهه الكريم، ويجعله في ميزان حسناتنا، وأن ينفع به الطلاب والدارسين.

و الله الموفق

المؤلفان الاغواط – الجزائر- ٢٠١٤/١٠/٠٧

١- المجموعات وعلاقتها بمفهوم الاحتمال:

يمكن إعطاء مفهوم الاحتمال من خلال أعمال بنود القيام ولا نعرف مدى صحتها ولا نعرف إذا كنا نستطيع القيام بها أم لا، وكذلك فإن تلك الأعمال تخضع لدرجة عدم التأكد، العلم الذي يبحث في دراسة حالة عدم التأكد هو نظرية الاحتمال، قبل الدخول في تعريف قانون الاحتمال، وجب ذكر بعض التعاريف التي ترافق دراسة مفهوم الاحتمال ومنها نظرية المجموعات.

١-١- نظرية المجموعات La théorie des ensembles:

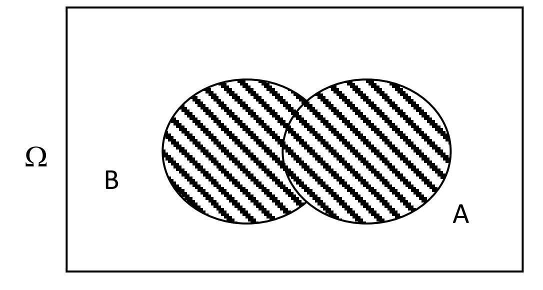
تسمى أي قائمة أو تجمع الأشياء بمجموعة، شريطة أن تكون القائمة معرفة تعريفا جيدا، وتسمى الأشياء المكونة لهذه المجموعات بعناصر (Les invidious) ونكتب $m \in A$ إذا كانت m عنصرا من المجموعة A، وسنفرض دائما ما لم ينص على ذلك صراحة أن جميع المجموعات التي ندرسها هي مجموعات جزئية من مجموعة ثابتة تسمى المجموعة الكلية (شاملة) ونرمز لها بالرمز (Ω, S, U) (L'ensemble fondamentale)، وللتعبير عن المجموعة الخالية (L'ensemble vide) والتي لا تحتوي على أي عنصر.

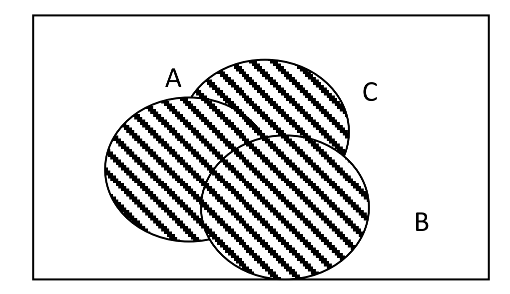
١-٢- عمليات المجموعات:

نفرض أن B,A مجموعتان اختياريتان، نعرف اتحاد A و B ويرمز له A∪B، بأنه مجموعة العناصر التي تنتمي إلى A أو B

$A \cup B = \{x : x \in A$ أو $x \in B\}$

ويمكن توضيح هذا في مخطط فن venn diagram.



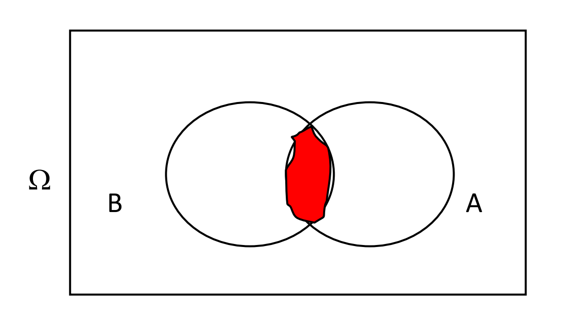


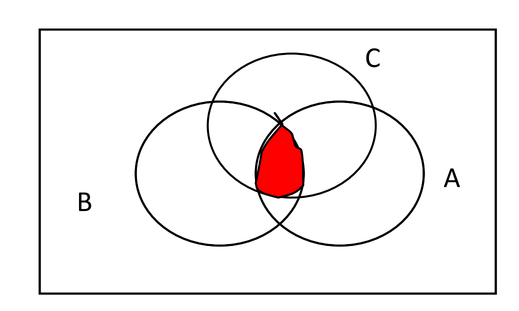
$A \cup B$ $\qquad$ $A \cup B \cup C$

ويعرف تقاطع A و B ويرمز له بالرمز $A \cap B$ بأنه مجموعة العناصر التي تنتمي إلى كل من A و B

$A \cap B = \{x : x \in A$ و $x \in B\}$

ويمكن توضيح هذا في مخطط فن:

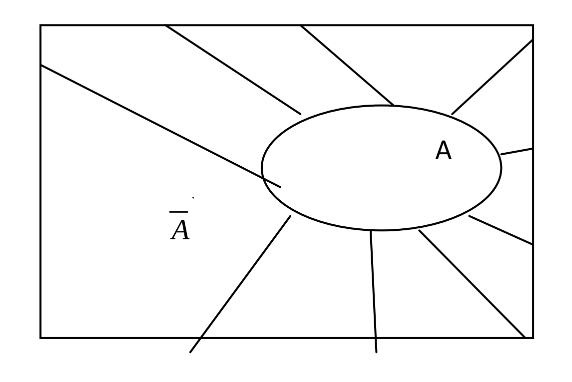


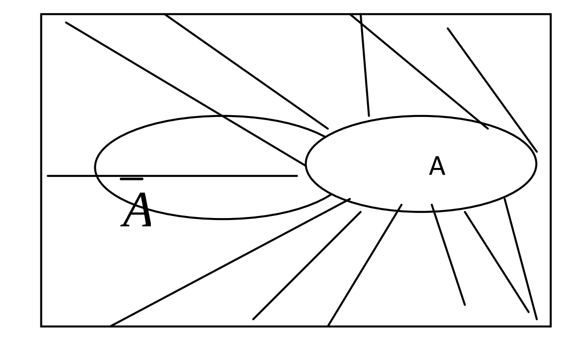


$A \cap B$ $\quad\quad$ $A \cap B \cap C$

وإذا كان $A \cap B = \phi$ أي لا يوجد عنصر مشترك في كل من A و B فيقال أن A و B منفصلتان، ويعرف المكمل Le complémentaire ويرمز له بالرمز $\overline{A}$ أو A^c نفيA(non A) بأنه مجموعة العناصر التي تنتمي إلى المجموعة الكلية ولا تنتمي إلى A

$A^c = \{x : x \in \Omega$ و $x \notin A\}$ $\quad\quad$ A^c مظلل





وسنلخص جبر المجموعات في الجدول التالي:

قوانين الثبات	$A \cup A = A$; $A \cap A = A$
قوانين الإدماج	$(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$
قوانين التبادل	$A \cup B = B \cup A$, $A \cap B = B \cap A$
قوانين التوزيع	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
قوانين الوحدة	$A \cup \varnothing = A$, $A \cap \Omega = A$ $A \cup \Omega = \Omega$, $A \cap \varnothing = \varnothing$
قوانين الاتمام	$A \cup \overline{A} = \Omega$, $A \cap \overline{A} = \varnothing$ $(A^c)^c = A$, $(\Omega)^c = \varnothing$, $\varnothing^c = \Omega$
قوانين دي مورغان	$\overline{A \cup B} = \overline{A} \cap \overline{B}$, , $\overline{A \cap B} = \overline{A} \cup \overline{B}$

مثال:

نفرض أن S مجموعة شاملة بحيث أن : S={1, 2,........10} و أن

C={5, 6,7,8,9} , B={2, 4,6,8,10} , A={1, 2,3,4,5}

أوجد: $\overline{A}, \overline{B}, \overline{C}, A \cap B, \overline{A \cap B}, A \cup B, A / C$

حل المثال:

$\overline{A}$={6, 7,8,9,10}، $\overline{B}$={1, 3,5,7,9} , $\overline{C}$={1,2, 3,4,5,10}

$\overline{A \cap B} = \Omega - A \cap B = \{1,3, 5,6,7,8,9 ,10\}$, $A \cap B = \{2,4\}$

أو نستعمل قانون دي مورغان

$$\overline{A \cap B} = \overline{A} \cup \overline{B} = \{1,3,5,6,7,8,9,10\}$$

$$A / C = A \cap \overline{C} = \{1,2,3,4\}$$

٢- مبادئ التحليل التوليفي L'analyse Combinatoire:

عند دراسة الاحتمالات لتجارب ذات فضاء منتهي، يلزمنا معرفة عدد نقاط فراغ (فضاء) الحوادث الأولية، وعدد نقاط الحوادث التي نرغب في معرفة احتمالاتها، وتبدو هذه المسألة صعبة ومعقدة ولها نلجأ عادة

للاستعيان بالقواعد الأساسية في تحليل المجموعات أو ما يسمى بالتحليل التوليفي.

سنتطرق في هذا الفصل إلى الموضوعات التالية:

- التباديل Permutations
- التراتيب Arrangements
- التوافيق Combinaisons

وقبل التطرق إلى هذه المواضيع الثلاثة لابد من الإشارة إلى نقطة أساسية تشكل المرتكز الحقيقي للتحليل التوليفي وهي قاعدة الضرب، أو ما يسمى بالمبدأ الأساسي في العد Dénombrement.

<u>٢-١- المبدأ الأساسي للعد Dénombrement:</u>

يعتمد هذا المبدأ على أنه إذا أمكن القيام بعمل ما بـ n_1 طريقة مختلفة، وإذا أمكن القيام بعمل بـ n_2 طريقة مختلفة فيمكن القيام بالعمليتين معا في آن واحد يعدد من الطرق مساويا إلى $n_2 x n_1$ طريقة ممكنة، ويمكن التعميم لهذا المبدأ على أكثر من عاملين، ويمكننا توضيح هذا المفهوم بالأمثلة التالية :

مثال: ما هي قدرة شبكة الهاتف النقال في الجزائر في منح عدد الخطوط للزبائن (حالة متعامل واحد).

حل المثال: وجب أخذ على الشكل لتسهيل الفهم 0X XX XX XX XX ، نلاحظ أن الصفر والخانة الموالية ثابتين والخانة الثالثة تتنافس عليها تسعة أرقام (1,........9) الخانة الرابعة إلى الخانة العاشرة (تتنافس عليها ١٠ أرقام (0,......9).

وبالتالي فعدد الخطوط هو 1.1.9.10.10.10.10.10.10.10

ملاحظة: قاعدة الضرب تطبق على الحالات التي تكون فيها الحوادث غير متنافية بالتبادل (أي وقوع حادث ما لا يمنع وقوع حادث آخر).

٢-٢- التباديل Permutations:

سنميز حالتين: التبادل دون تكرار العناصر Permutation sans répétition.

التباديل مع تكرار العناصر Permutation avec répétition

٢-٢-١- التباديل دون تكرار العناصر: نفترض أن لدينا من المفردات يراد ترتيبها أو وضعها في n عن الأماكن أو المواضع، بحيث أن كل

مفردة تستعمل ولمرة واحدة فقط، فإن عدد الطرق الممكنة للترتيب هنا يسمى مضروب n أو n عاملي (n factorielle) بحيث أن $n \in \mathbb{N}$

مثال: يوجد عندنا ٣كتب C,B,A، فعملية التبديل تكون على النحو التالي: $P_n = 3! = 6$ (يوجد لدينا ٦ طرق): ABC, ACB, BAC, BCA, CAB, CBA

ملاحظة: حالة التباديل الدائرية $P_{\bar{n}} = (n-1)!$

مثال: كل طريقة يمكن لـ٥ أشخاص أن يجلسوا فوق طاولة مستديرة.

$$P_{\bar{5}} = (5-1)! = 24$$

٢-٢-٢- التباديل مع التكرار: فنستخدم القانون التالي:

$$P_n^{n_1, n_2, \ldots\ldots n_K} = \frac{n!}{n_1!.n_2!\ldots\ldots.n_K!}$$

مثال: ما هو عدد التباديل المختلفة التي يمكن تكوينها من جميع أحرف كلمة recherche

حل المثال:

$r = 2, e = 3, c = 2, h = 2$

$$P_9^{2,2,2,3} = \frac{9!}{2!.2!.2!.3!} = 7560$$

٢-٣- الترتيب : Arrangements

ترتيبة لـk عنصر من بين n عنصرا هي تعرف ترتيب على k عنصر مختارة من بين n عنصر ونرمز له بـ A_n^k ، لعدد الترتيبات لـk عنصر من بين n عنصر $A_n^k = \frac{n!}{(n-k)!}$

٢-٤- التوافيق: Combinaisons

نهتم في التراتيب برتبة العناصر، وهكذا فإن ABC يختلف عن BCA ولكننا في العديد من المسائل نهتم فقط بانتقاء أو اختيار العناصر دون النظر إلى الترتيب، ندعو بمثل هذه الاختيارات بالتوافيق فمثلا ABC و BAC التوفيق نفسه تسمى توافيق n عنصر مأخوذ: r منها في كل مرة بـ C_n^r بحيث أن $C_n^r = \frac{n!}{r!(n-r)!}$

٢-٤-١- خصائص التوفيقات:

$$1)\ C_n^r = C_n^{n-r} \qquad 2)\ C_n^r = C_{n-1}^r + C_{n-1}^{r-1}$$

$$3)\ C_n^n = \frac{n!}{0!\,n!} = 1 \qquad 4)\ C_{n+1}^r = C_n^{r-1} + C_n^r$$

٢-٤-٢- نظرية ثنائي الحد: ليكن a و b عددين كيفين، لدينا :

$$(a+b)^n = \sum_{r=0}^{n} C_n^r a^{n-r} b^r$$

مثال ١: $(a+b)^3 = a^3 + 3a^2b + b^3$

مثال٢: الصيغة التالية تعرف بصيغة باسكال La formule de pascale

، المطلوب إثباتها : $C_{n+1}^{r} = C_n^{r-1} + C_n^r$

حل مثال ٢: $C_n^{r-1} + C_n^r = \frac{n!}{(r-1)!(n-r+1)!} + \frac{n!}{r!(n-r)!}$

لكي نتحصل على نفس المقام نضرب الأول في $\frac{r}{r}$ والثاني في $\frac{n-r+1}{n-r+1}$

لأن

$$\frac{rn!}{r!(r-1)!(n-r+1)!} + \frac{n!(n-r+1)}{r!(n-r+1)(n-r)!} = \frac{n!(r+n-r+1)}{r!(n-r+1)(n-r)!}$$

$$= \frac{(n+1)n!}{r!(n-r+1)!} = \frac{(n+1)!}{r!(n-r+1)!} = C_{n+1}^r$$

$$(n-r+1)! = (n-r+1)(n-r)! \ \text{،} \ r! = r(n-1)!$$

تمارين: نظرية المجموعات والتحليل التوليفي:

تمرين ٠١: لتكن C,B,A أحداث كيفية، برهن أن :

$$A \cup B \cup C = A+B+C - A\cap B - A\cap C - B\cap C + A\cap B\cap C$$

تمرين ٠٢: في عملية سبر للآراء لـ ١٥٠ عامل حول تولي ٣ مرشحين لمنصب رئاسة الاتحاد النقابي التابع للمؤسسة "س"، حيث تبين ما يلي:
٦٥ عامل يؤيد المرشح A، ٥٠ يؤيد المرشح B، ٥٠ يؤيد المرشح C، ٢٥ يؤيدون A و B، ٢٥ يؤيدون A و C، ٢٠ يؤيدون B وC ، و ٤٠ يؤيدون هؤلاء المشتركين إطلاقا.

المطلوب: ما هو عدد المؤيدين للمرشحين الثلاثة؟

تمرين ٠٣: لدى أحمد ٣ كتب في الرياضيات و٥ كتب في الفقه و ٤ كتب في الفيزياء ما هو عدد الطرق الممكنة لوضع هذه الكتب على مرفع (رف) وذلك في الحالات التالية.

أ- توضع الكتب المتعلقة بنفس الموضوع متجاورة؟

ب- كتب الفقه فقط توضع متجاورة؟

تمرين ٠٤: بفرض عدم السماح بالتكرار أحسب:

أ- كم عدد مكونا من ثلاثة أرقام يمكن تركيبه من الأرقام التالية ٢، ٣، ٤، ١، ٩.

ب- كم عدد منهم أقل من ٥٠٠؟

ج- كم عدد منهم زوجيا ؟ وكم عدد منهم فرديا؟

د- كم عدد منهم مضاعف للعدد ٥؟

<u>تمرين ٠٥:</u> عين الأعداد الطبيعية n التي تحقق الشرط المعطى:

$$1)\, C_n^4 - C_n^3 = n^3 - 3n^2 + 2n,,$$

$$2)\, A_n^2 = 42$$

$$3)\, 4A_n^2 + 252 = A_{3n}^2$$

٣- مبادئ الحساب الاحتمالي :

معنى الاحتمال: ببساطة الاحتمال هو عبارة عن تعميم لمبدأ النسبة فإذا رمزنا للحدث بـA، فإن احتمال وقوعه هو A/n.

٣-١- مفاهيم أساسية:

٣-١-١- التجربة: هي العملية التي نحصل منها على النتائج مثل تجربة رمي حجر نرد، رمي قطعة نقود...

٣-١-٢- فضاء العينة: هو عبارة عن جميع النتائج الممكن الحصول عليها من التجربة، ويرمز لها بـΩ ، مثلا تجربة رمي قطعين نقد منتظمتين فضاء العينة يكون كالتالي: {HH,HT,TH,TT} = و، حيث H يرمز للصورة، T يرمز لظهور الكتابة.

٣-١-٣- الحدث: ويمثل مجموعة جزئية من عناصر فضاء العينة ويرمز للحوادث بالحروف C,B,A .

مثلا في التجربة السابقة نعرف ، A الحدث الذي يهتم بظهور صورة واحدة على الأقل فيكون {HT, TH, HH} =A.

٣-٢- بديهيات الاحتمال:

البديهية ١: $0 \leq p(A) \leq 1$

البديهية ٢: $p(\Omega) = 1$

البديهية ٣: $p(A_1 \cup A_2 \cup \ldots\ldots \cup A_i \cup \ldots) = P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

حيث A_1, A_2...أحداث غير متصلة، وتدعى هذه البديهيات الثلاثة ببديهيات كولموغوروف (Axiomes de Kolmogorov).

٣-٣- طرق قياس الاحتمال: هناك ثلاثة طرق لقياس الاحتمال وهي الطريقة التقليدية (اللابلاسية)، الطريقة الإحصائية، الطريقة الهندسية وسنتناول الطريقة الأولى والتي تنص على : $p(A) = \frac{n.f.deA}{n.t.de\Omega}$ ، حيث ($n.f.deA$)عدد الحالات الملائمة لA و ($n.t.de\Omega$)عدد الحالات الملائمة ل Ω .

٣-٤- الاحتمال الشرطي La probabilité conditionnelle:

يعرف الاحتمال الشرطي على أنه احتمال وقوع الحدث A مثلا، مشروطا بحدث آخر وليكن B قد وقع فعلا، ويرمز لهذا الامتثال بالرمز P(A/B) ويعرف كالتالي:

$$P(A/B)=\frac{P(A\cap B)}{P(B)};;P(B)>0$$

نفس الأسلوب مع(B/A) P حيث:

$$P(B/A)=\frac{P(A\cap B)}{P(A)};;P(A)>0$$

٣-٥- الأحداث المستقلة Les événements indépendants:

يقال بأن الحدثين A و B مستقلان إذا كان حصول أحدهما لا يؤثر بحصول الآخر أي:

$P(B/A) = P(B)$

$P (A/B) = P(A)$

$P (A\cap B)= P(A). P(B)$

$P (A\cap B\cap C) = P(A). P(B). P(C)$

مثال:

نرمي قطعة نقدية ٣ مرات ونفرض أن الأحداث الأولية متساوية الاحتمال فيكون فضاء العينة كما يلي:

Ω= {HHH, HHT...............................TTT}

نعتبر الأحداث:

A: نحصل على وجه في الرمية الأولى؟

B: نحصل على وجه في الرمية الثانية؟

C: نحصل على وجه خلال رميتين متواليتين لا غير؟

هل أن A وB مستقلان؟

هل أن A وC مستقلان؟

هل أن B وC مستقلان؟

حل المثال:

$$A=\{HHH,HHT,HTH,HTT\},P(A)=\frac{1}{2};;B=\{HHH,HHT,THH,THT\},P(B)=\frac{1}{2}$$
$$C=\{HHT,THH\},P(C)=\frac{2}{8}=\frac{1}{4};;A\cap B=\{HHH,HHT\},P(A\cap B)=\frac{1}{4}$$
$$A\cap C=\{HHT\},P(A\cap C)=\frac{1}{8};;B\cap C=\{THH,HHT\},P(B\cap C)=\frac{1}{4}$$
$$P(A).P(B)=\frac{1}{2}.\frac{1}{2}=\frac{1}{4}=P(A\cap B)$$
$$P(A).P(C)=\frac{1}{2}.\frac{1}{4}=\frac{1}{8}=P(A\cap C)$$
$$P(B).P(C)=\frac{1}{2}.\frac{1}{4}=\frac{1}{8}\neq P(B\cap C)$$

إذن A و B مستقلان وA و C مستقلان و B و C غير مستقلان

٣-٦- نظرية بابيز La théorie de Bayes:

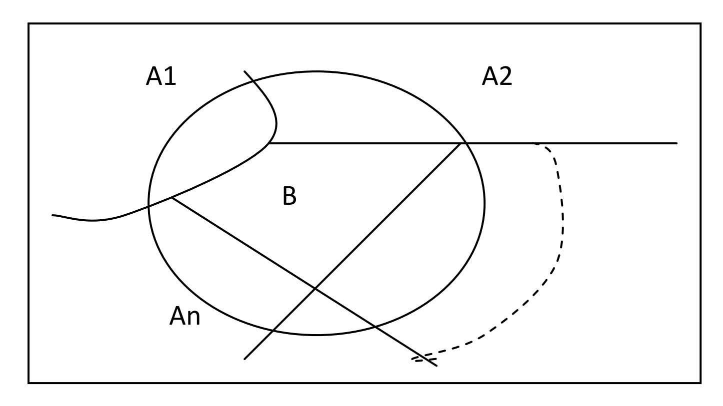
إذا كان Ω يمثل فضاء العينة تجربة ما، نفرض أن الأحداث A_1 ,A_2 ,

A_N......... لفضاء العينة Ω ، أي أن الأحداث A_i متنافية مثنى مثنى،

واتحادهما هو Ω، ونفرض أن B حدث ما، يمكن كتابة:

$B=\Omega \cap B = A_1 \cup A_2 \cup \ldots\ldots\ldots \cup A_n) \cap B$

$= (A_1 \cap B) \cup (A_2 \cap B) \cup \ldots.(A_n \cap B)$



نظرية بابيز تعرف كما يلي:

$$P(A_i / B) = \frac{P(A_i).P(B / A_i)}{P(A_1).P(B / A_1) + ... + P(A_n).P(B / A_n)}$$

هذه العلاقة لها أهمية خاصة لكون باستعمالها كأساس في الإحصائيات "البيزية".

تمارين: الحساب الاحتمالي:

تمرين١: نفرض أن ألفينا قطعة نقود وحجر نرد معا كون فضاء العينة Ω ، ثم عبر بوضوح عن الأحداث التالية:

أ) ظهور الصورة مع عدد فردي؟

ب)ظهور عدد أولي؟

ج) ظهور الكتابة مع عدد زوجي؟

أي من الأحداث C,B,A يتنافى مع الآخر؟

تمرين٢: نفرض أن $\Omega=\{a_1,a_2,a_3,a_4,a_5\}$ وأن p دالة احتمال معرفة على Ω .

أ- أوجد $p(a_3)$ إذا كان $p(a_4)=p(a_5)=\frac{1}{8}$ ، و $p(a_1)=p(a_2)=\frac{1}{6}$.

ب- أوجد $p(a_1),p(a_3),p(a_5)$ ، إذا كان $p(a_2)=p(a_4)=\frac{3}{15}$ و $p(a_3)=2p(a_5)$ و $p(a_1)=6p(a_5)$

تمرين٣: إذا اخترنا عددا عشوائيات x بين ١و ١٠٠، فما هو احتمال أن يكون العددx قابل للقسمة على ٣ أو ٤؟

تمرين٤: ألقينا حجري نرد ، أوجد P) احتمال أن يكون:

أ)- المجموع ٦، ٢) أن يظهر العدد١، ٣) المجموع ٤ أو أقل.

ب) إذا كان العددان الظاهران مختلفان، نفس المطلوب السؤال السابق.

تمرين٥: في كلية العلوم الاجتماعية رسب ١٠% من الطلبة في امتحان الإحصاء و ١٠% من الطلبة في امتحان المنهجية ورسب ٥% في امتحان الإحصاء والمنهجية، اختير أحد الطلبة بطريقة عشوائية وتبين ما يلي:

أ- إذا كان راسبا في الإحصاء فما هو احتمال أن يكون راسبا في المنهجية؟

ب- إذا كان راسبا في المنهجية فما هو احتمال أن يكون راسبا في الإحصاء؟

ج- ما هو احتمال أن يكون راسبا في مقياس الإحصاء أو المنهجية؟

٤- المتغيرات العشوائية والتوزيعات الاحتمالية Les variables aléatoires et distributions probabilités

تعريف: هو الدالة التي قيمتها رقم حقيقي، يتم تحديده بواسطة عنصر من فضاء العينة.

مثال١: لتكن التجربة هي رمي ثلاث قطع نقود، وليكن x المتغير العشوائي الذي يمثل ظهور الوجه H، فالمتغير العشوائي x يأخذ القيم ٠ أو ١ أو ٢ أو ٣

فضاء الأحداث الابتدائية	HHH	TTH	THT	HTT	THH	HTH	HHT	TTT
قيمة x	٣	١	١	١	٢	٢	٢	٠

٤-1- تصنيف المتغيرات العشوائية:

٤-1-١- المتغير العشوائي المتقطع: نقول أن المتغير العشوائي متقطع إذا كانت مجموعة قيمه الممكنة منتهية أولا نهائية قابلة للعد.

٤-1-٢- المتغير العشوائي المستمر: نقول أن المتغير العشوائي مستمر إذا كانت مجموعة قيمه الممكنة لا نهائية وغير قابلة للعد.

٤-2- المتغيرات العشوائية المتقطعة وتوزيعاتها الاحتمالية:

إذا كان x متغيرا عشوائيا على فضاء الأحداث الابتدائية Ω و كانت مجموعة القيم التي يأخذها هي: $R_x = \{x_1, x_2 ....x_n ...\}$، وإذا عرفنا الدالة $f_{(x)}$ على الشكل $f(x_i) = P(x = x_i)$ ، فإننا نلاحظ أن الدالة $f(x)$ تحقق الشرطين:

$$1)\ f(x_i) \geq 0$$
$$2)\sum_i f(x_i) = 1$$

مثال: خمس بطاقات تحمل الأرقام ٥، ٤، ٣، ٢، ١، نسحب ثلاث بطاقات منها بشكل عشوائي، ونرمز بـx المتغير العشوائي الدال على مجموع أرقام البطاقات المسحوبة، والمطلوب كتابة جدول التوزيع الاحتماليx.

حل المثال: نلاحظ أن فضاء الأحداث الابتدائية لهذه التجربة هو:

Ω= {(1,2,3) , (1,2,4), (1,2,5), (1,3,4), (1,3,5), (1,4,5), (2,3,4), (2,3,5), (2,4,5), (3,4,5) }

ويأخذ المتغير العشوائيx القيم التالية: R_x ={6,7,8,9,10,11,12}

إن الحدث [x=٦] يوافق الحدث {(1,2,3)} ويكون $p(6)=\frac{1}{6}$ وهكذا دواليك، ويكون التوزيع الاحتمالي كما يلي:

X	٦	٧	٨	٩	10	11	12
$f_{(x)}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

أما دالة التوزيع التراكمية F للمتغير x (متغير عشوائي متقطعا وتوزيعه هو f) هي كما يلي: $F(x)=\sum_{x_i<x} f(x_i)$، ودالة التوزيع الاحتمالي $F_{(x)}$ تحقق الخواص التالية:

١)− $F_{(x)}$ دالة غير متناقصة ومستمرة على الأقل من اليسار عند كل نقطة x من R.

٢) $\lim_{x\to+\infty} F(x)=1,;\lim_{x\to-\infty} F(x)=0$

٣) $P(a\le x\le b)=F(b)-F(a)$

<u>مثال:</u> بالعودة إلى المثال السابق يمكن أن نعرض دالة التوزيع المجتمع كما يلي:

ومن أجل كل قيمة $x>12$، فإن $F_{(x)} = 1$

X	٦	٧	٨	٩	10	11	12	١٣
$F_{(x)}$	٠	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{4}{10}$	$\frac{6}{10}$	$\frac{8}{10}$	$\frac{9}{10}$	١

٤-3- المتغيرات العشوائية المستمرة وتوزيعاتها الاحتمالية:

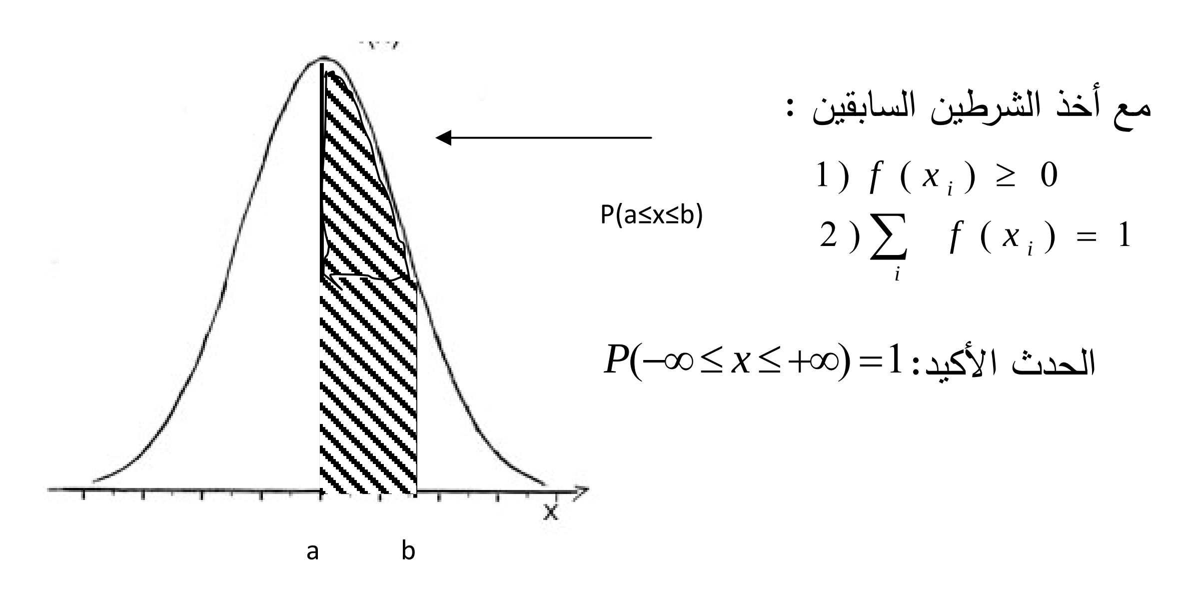
يأخذ مجالا معينا من مجموعة الأعداد الحقيقية IR، لذلك سيمثل توزيعه الاحتمالي دالة مستمرة تسمى بدالة الكثافة الاحتمالية La fonction de densité probabilité، والتي يرمز لها بـ $f_{(x)}$، والاحتمال في المتغيرات العشوائية المستمرة هو عبارة عن مساحة تحت المنحنى، وبالتالي يكون الاحتمال إحدى الصور الآتية

$$P(a \le x \le b), P(x \le a), \quad P(x \ge b)$$

والتي يمكن إيجادها جميعا باستخدام التكاملي:

$$P(x \le a) = \int_{-\infty}^{a} f(x)\,dx, \quad P(x \ge b) = \int_{b}^{+\infty} f(x)\,dx,$$

$$P(a \le x \le b) = \int_{a}^{b} f(x)\,dx,$$



مع أخذ الشرطين السابقين :

$1)\ f(x_i) \geq 0$

$2)\ \sum_i f(x_i) = 1$

الحدث الأكيد: $P(-\infty \leq x \leq +\infty) = 1$

<u>مثال:</u> ليكن x متغيرا عشوائيا مستمرا معرفا بالدالة الاحتمالية التالية:

$$f(x) = \begin{cases} cx^3 ; 0 \leq x \leq 4 \\ 0 / w \end{cases}$$

أ) أوجد قيمة C، حتى تكون $f_{(x)}$ دالة كثافة احتمالية.

ب) أوجد $P(0 \leq x \leq 2)$

<u>حل المثال:</u> بما أن $f_{(x)}$ تحقق الخاصية ١ إذا كان $C \geq 0$، فإنها تحقق الخاصية ٢ كي تكون دالة كثافة.

$$1)\int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{4} cx^3 dx = 1 \Leftrightarrow 64c = 1 \Rightarrow c = \frac{1}{64}$$

$$2)P(0 \le x \le 2) = \int_{0}^{2} \frac{1}{64} x^3 dx = \frac{1}{64}$$

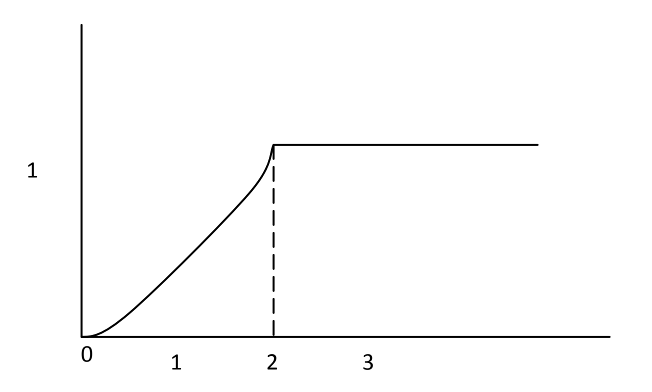
<u>دالة التوزيع التراكمية:</u> إذا كان x متغيرا عشوائيا مستمرا وتوزيعه هو f فإن $F(x) = \int_{-\infty}^{x} f(t)dt,$

<u>مثال:</u> نفرض أن x متغير عشوائي مستمر له دالة التوزيع التالية:

$$f(x) = \begin{cases} \frac{1}{2}x; 0 \le x \le 2 \\ 0 / w \end{cases}$$

وتكون دالة التوزيع المتراكمة F كما يلي:

$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ \frac{1}{4}x^2 & si \quad 0 \le x \le 2 \\ 1 & si \quad x > 2 \end{cases}$$



٤-4- التوزيعات المشتركة للمتغيرات العشوائية:

نحتاج في كثير من الحالات عند وصف تجربة لأكثر من متغير عشوائي فعندما نلقي حجري نرد يمكن أن نرمز لمجموع رقمي الوجهين الظاهرين بـ x، ونرمز للقيمة العظمى لهذين الرقمين بــy، فمجموعة نتائج هذه التجربة يمكن أن ندل عليها بواسطة الثنائيات (x,y).

٤-4-١- الأشعة العشوائية الثنائية المتقطعة:

تعريف: نقول أن الشعاع العشوائي (x,y) متقطع إذا كانت مجموع القيم التي يأخذها مجموعة منتهية أو غير قابلة للعد. وسنوضح ذلك في الجدول التالي:

y / X	y1	y2	....	Y j	...	المجموع
x1	$f(x_1, y_1)$	$f(x_1, y_2)$	$\cdots$	$f(x_1, y_j)$	$\cdots$	$\sum_{j\geq1} f(x_1, y_j)$
x2	$f(x_2, y_1)$	$f(x_2, y_2)$	$\cdots$	$f(x_2, y_j)$	$\cdots$	$\sum_{j\geq1} f(x_2, y_j)$
/	/	/	/	/	$\cdots$	/
Xi	$f(x_i, y_1)$	$f(x_i, y_2)$	$\cdots$	$f(x_i, y_j)$	$\cdots$	$\sum_{j\geq1} f(x_i, y_j)$
/	/	/	/	/	/	/
المجموع	$\sum_{i\geq1} f(x_i, y_1)$	$\sum_{i\geq1} f(x_i, y_1)$	$\cdots$	$\sum_{i\geq1} f(x_i, y_j)$	$\cdots$	١

مثال: نفرض أن x و y متغيران عشوائيان لهما جدول التوزيع المشترك.

y / x	2	3	4	المجموع
1	٠,١٥	٠,٢	٠,٢٥	٠,٦

٠,٤	٠,٠٥	٠,١	٠,٢٥	2
١	٠,٣	٠,٣	٠,٤	المجموع

التوزيع الهامشي لـx

$$L(x) = \begin{pmatrix} x_1 & x_2 \\ 0.6 & 0.4 \end{pmatrix}$$

التوزيع الهامشي لـ y

$$L(y) = \begin{pmatrix} y_1 & y_2 & y_3 \\ 0.4 & 0.3 & 0.3 \end{pmatrix}$$

٤-4-٢- التوزيع الاحتمالي المشترك والهامشي للمتغير الكمي المستمر:

تعريف: الدالة الاحتمالية لمتغيرين عشوائيين x وy والتي تسمى بدالة الكثافة المشتركة لـx و y بـ

$$1)\ f(x, y) \geq 0$$

$$2)\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x, y)\,dx\,dy = 1$$

ونعرف دالة كثافة الاحتمال الهامشية للمتغير x كما يلي:

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\,dy,$$

ودالة كثافة الاحتمال الهامشية للمتغير y كما يلي

$f(y) = \int_{-\infty}^{\infty} f(x,y)\,dx,$ ، أما دالة كثافة الاحتمال

الشرطية للمتغير x فهي: $f(x/y) = \dfrac{f(x,y)}{f(y)}$ ، ودالة

كثافة الاحتمال الشرطية للمتغير y هي

$f(y/x) = \dfrac{f(x,y)}{f(x)}$.

<u>مثال:</u> نفرض أن دالة الكثافة المشتركة للمتغيرين العشوائيين x و y

تعطى بـ $f(x,y) = \begin{cases} cxy; 0 < x < 3; 1 < y < 2 \\ 0/w \end{cases}$

١) أوجد الثابت C.

٢) أوجد دالة كثافة الاحتمال الهامشية للمتغير x ؟

٣) أوجد دالة كثافة الاحتمال الهامشية للمتغير y؟

٤) أوجد دالة كثافة الاحتمال الشرطي للمتغير x ؟

٥) أوجد دالة كثافة الاحتمال الشرطي للمتغير y ؟

ماذا نستنتج ؟

حل المثال: يجب أن يكون لدينا مجموع الاحتمالات مساويا لــ١ أي ان $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = 1$ ، وباستخدام تعريف $f(x,y)$ تصبح قيمة التكامل

$$\int_0^3\int_1^2 cxydxdy = c\int_0^3\left[\int_1^2 xydy\right]dx = c\int_0^3\left[\frac{xy^2}{2}\right]_1^2 dx = c\int_0^3\left(\frac{4x}{2}-\frac{x}{2}\right)dx$$

$$= c\int_0^2\frac{3}{2}xdx = \frac{3}{2}c\left[\frac{x^2}{2}\right]_0^3 = \frac{27c}{4} = 1 \Rightarrow c = \frac{4}{27}$$

٢) دالة كثافة الاحتمال الهامشي للمتغير x

$$f(x) = \int_1^2 \frac{4}{27}xydy = \frac{2}{9}x$$

٣) دالة كثافة الاحتمال الهامشي للمتغير y

$$f(y) = \int_0^3 \frac{4}{27}xydy = \frac{2}{3}y$$

٤) دالة كثافة الاحتمال الشرطي للمتغير x

$$f(x/y) = \frac{f(x,y)}{f(y)} = \frac{\frac{4}{27}xy}{\frac{2}{3}y} = \frac{2}{9}x$$

٥) دالة كافة الاحتمال الشرطي للمتغير y

$$f(x/y) = \frac{f(x,y)}{f(y)} = \frac{\frac{4}{27}xy}{\frac{2}{9}x} = \frac{2}{3}y$$

نستنتج أن x و y مستقلين لأن $f(x,y) = f(x).f(y)$

٤-5- بعض القيم المميز : للتوزيعات الاحتمالية:

٤-5-١- التوقع الرياضي L'espérance mathématique

(حالة ت.متقطع):

تعريف١: ت.ر.م.ع. متقطع x دالة الاحتمالية $f(x)$ هو المقدار $E(x) = \sum_x xf(x_i)$ وعادة ما يرمز للتوقع بــμ.

تعريف٢: ت.ر.م.ع مستمر x دالة كثافته الاحتمالية $f(x)$ هو المقدار $E(x) = \int_{-\infty}^{+\infty} xf(x)dx$

أما خواص التوقع الرياضي فنوجزها في النقاط التالية:

(P_1) : E (x+y) = E (x) + E (y)

(P_2) : E (ax) = a E (x), $\forall a \in$

(P3): Si x est un caractère constant, tel que

$\forall \omega \in \Omega, x(\omega) = k$, alors : E(x) = K

<u>مثال:</u> لدينا التوزيع المشترك: أحسب ما يلي E(x), E (y), E(xy), E(x+y)

Y / x	2	3	4	L(x)
1	٠,١	٠,١	٠,٢	٠,٤
2	٠,٣	٠,٢	٠,١	٠,٦
L(y)	٠,٤	٠,٣	٠,٣	١

<u>حل المثال:</u>

y_i	٢	٣	٤	المجموع
$P(y_i)$	٠,٤	٠,٣	٠,٣	١

X_i	١	٢	المجموع
$P(x_i)$	٠,٤	٠,٦	١

$$E(x)=\sum_{x} x f(x_i)=1.0,4+2.0,6=1,6$$

$$E(y)=\sum_{y} y f(y_i)=2.0,4+3.0,3+4.0,3=2,9$$

$$E(x+y)=\sum_{i}\sum_{j}(x+y) f(x,y)=(1+2).0,1+(1+3).0,1+.....+(2+4).0,1=4,5$$

$$E(xy)=\sum_{i}\sum_{j}(x_i y_j) h(x_i,y_j)=2(0,1)+3(0,1)+....+8(0,1)=4,5$$

أو بطريقة أخرى لحساب $E(x+y)$

$$E(x+y)=E(x)+E(y)=1,6+2,9=4,5$$

٤-5-٢- التباين (حالة م.ع.م) La variance:

تعريف: ليكن x متغير عشوائي توقعه الرياضي μ ، إن العدد $E(x-\mu)^2$ ، يسمى بتباين المتغير x ، وسوف نرمز له بـ $v(x)$ ونسمى جذر $\sigma(x)$ بالانحراف المعياري Ecart type.

$$v(x)=E((x-\mu)^2)=\sum_{i}(x-\mu)^2 P(x)$$

X_i	١	٢	٣	٤	٥	٦
$p(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$(x-\mu)^2$	$\frac{25}{4}$	$\frac{9}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{9}{4}$	$\frac{25}{4}$

<u>مثال:</u> في تجربة رمي حجر نرد، أحسب التوقع $E(x)$ والتباين $v(x)$ ؟

حل المثال :

$$\mu = E(x) = \sum x_i p(x_i) = \frac{21}{6} ،$$

$$v(x) = \frac{1}{6}\left(\frac{25}{4} + \frac{9}{4} + \frac{1}{4} + \frac{1}{4} + \frac{9}{4} + \frac{25}{4}\right) = \frac{35}{12} ،$$

$$\sigma(x) = \sqrt{\frac{35}{12}} = 1,70$$

هناك صيغة مختصرة لحساب التباين تدعى بصيغة Huyghens

وهي $v(x) = E(x^2) - \mu^2$ أما خصائص التباين فنوجزها في النقاط التالية:

1) V (a) = 0

2) $V(ax) = a^2 V(x)$, 3) $V(ax + b) = a^2 V(x)$

الإثبات يترك للطالب (باستخدام تعريف التباين).

٤-5-٣- التوقع الرياضي (حالة م.ع. مستمر):

تعريف: ليكن x متغير عشوائي له دالة كثافة $f(x)$ فإن توقعه الرياضي يكون كما يلي: $E(x) = \int_{-\infty}^{+\infty} x f(x) dx$

مثال١: إذا كان المتغير المستمر x له دالة كثافة احتمال :

$$f(x) = \begin{cases} \frac{1}{64} x^3 ; 0 \leq x \leq 4 \\ 0 / w \end{cases}$$

حل مثال١: $E(x) = \int_0^4 \frac{1}{64} x^4 dx = 3,2$

مثال٢: في حالة متغيرين عشوائيين x و y ، تعطى دالة الكثافة المشتركة لمتغيرين عشوائيين x و y بـــ

$$f(x, y) = \begin{cases} \frac{9}{992} x^2 y^2 ; 0 < x < 2 ; 1 < y < 5 \\ 0 / w \end{cases}$$

أوجد $E(x)$, $E(y)$, $E(xy)$, $E(3x+2y)$.

حل المثال٢:

$$E(x)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)dxdy=\int_{1}^{2}\int_{1}^{5}\frac{9}{992}x^3y^2dxdy=1,5$$

$$E(y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)dxdy=\int_{1}^{2}\int_{1}^{5}\frac{9}{992}x^2y^3dxdy=3,77$$

$$E(xy)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy=\int_{1}^{2}\int_{1}^{5}(\frac{9}{992}x^3y^3)dxdy=5,66$$

$$E(3x+2y)=\int_{1}^{2}\int_{1}^{5}\frac{9}{992}(3x^3y^2+2x^2y^3)dxdy=12,04$$

طريقة أخرى: $E(3x+2y)=3E(x)+2E(y)=3(1,5)+2(3,77)=12,04$

٤-5-٤- التباين (ح.م.ع مستمر):

__تعريف:__ ليكن x متغير عشوائي دالة كثافته الاحتمالية $f(x)$ ، التباين $\sigma^2(x)=v(x)=E((x-\mu)^2)$ ، يعرف كالتالي:

$$v(x)=\int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx$$

__مثال:__ عين التوقع الرياضي والانحراف المعياري للمتغير العشوائي x الذي كثافته الاحتمالية:

$$f(x) = \begin{cases} \frac{1}{5} & 1 \le x \le 6 \\ 0 / w \end{cases}$$

حل المثال:

$$\mu = E(x) = \int_{-\infty}^{\infty} xf(x)dx = \int_{1}^{6} \frac{x}{5} dx = \frac{35}{10}$$

$$v(x) = E(x^2) - \mu^2$$

$$v(x) = \int_{1}^{6} \frac{x^2}{5} dx - (\frac{35}{10})^2 = 2.08$$

ويكون الانحراف المعياري $\sigma(x) = \sqrt{v(x)} = \sqrt{2.08} = 1.44$

٤-6- التباين المشترك (التغاير) بين متغيرين عشوائيين La covariance:

تعريف: يرمز للتغاير بين المتغيرين العشوائيين x و y بـ cov (x,y)

ويعرف وفقا للعلاقة التالية: $\text{cov}(x, y) = E\left[(x - \mu_x)(y - \mu_y)\right]$ ....(I)

حيث μ_x يرمز للتوقع x، μ_y يرمز للتوقع y.

ويمكن كتابة الصيغة (I) على الشكل التالي:

$$\text{cov}(x, y) = E(xy) - \mu_x \mu_y$$

أما خواص التباين المشترك فنوجزها في النقاط التالية:

إذا كان y,x متغيرين عشوائيين وكان $a, b, c, d \in \mathbb{R}$

فإن:

1) cov (ax+b , C y+d) = ac cov (x.y)

2) V (x± y)= V (x) + V(y) ± 2 cov (x.y)

٤-7- الارتباط بين متغيرين عشوائيين La corrélation:

تعريف: يرمز للارتباط بين متغيرين عشوائيين x و y بـ $\rho(x,y)$ ويعرف كما يلي $\rho(x,y) = \frac{\operatorname{cov}(x,y)}{\sigma(x)\sigma(y)}$، أما خصائصه فنوجزها فيما يلي:

إذا كان x و y متغيرين عشوائيين وكان a,b,c,d أعداد ثابتة حقيقية فإن:

$$1)\rho(x,y) = \rho(y,x)$$
$$2)\rho(x,x) = 1, \quad \rho(x,-x) = -1$$
$$3)\forall a,c \neq 0, \quad \rho(ax+b, cy+d) = \rho(x,y)$$
$$4) -1 \leq \rho(x,y) \leq 1$$

٤-8- العزوم Les moments:

إن الفائدة من إيجاد العزوم هي معرفة شكل التوزيع الاحتمالي.

تعريف١: إذا كان $\mu_x = x^r$ حيث r صحيح موجب فإن المقدار $m_r = E[u(x)] = E(x^r)$ يدعى بالعزم الابتدائي من الرتبة r للمتغير العشوائيx ، وللعزوم أهمية خاصة في دراسة التوزيعات الاحتمالية، فمن أجل r=1 نجد $m_1 = E(x) = \mu$.

مثال: إذا كان x متغيرا عشوائيا جدول توزيعه الاحتمالي :

X	-٢	٤
$f(x)$	٠,٦	٠,٤

فإن

$$m_1 = E(x) = \sum xf(x)$$
$$m_1 = -2(0,6) + 4(0,4)$$
$$m_2 = E(x^2) = (-2^2)(0,6) + (4^2)(0,4) = 8,8$$

تعريف ٢: إذا كان x متغيرا عشوائيا وكان $u(t) = e^{tx}$ حيث t_متغير حقيقي فإن: $E[u(t)] = E(e^{tx})$ تدعى هذه بالدالة المولدة للعزوم (La fonction génératrice des moments) للمتغير x ونرمز بـ $M_x(t)$ وبالتالي: $M_x(t) = E(e^{tx})$

إذا كان X متغيرا عشوائيا متقطعا فإن

$$M_x(t) = \sum_{i \geq 1} e^{txi} p(x = x_i)$$

وإذا كان X متغيرا عشوائيا مستمرا فإن

$$M_x(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

٤-9- متباينتي ماركوف وتشيبيشيف- بينامية:

٤-9-١- متباينة ماركوف Inégalité de Markov:

إذا كان x متغير عشوائي يأخذ قيما موجبة، عندئذ من أجل كل عدد K>٠ يكون $p(x \geq k) \leq \frac{E(x)}{k}$

البرهان: سنذكر البرهان من أجل X متغيرا مستمرا كثافته $f(x)$

$$E(x) = \int_0^{\infty} xf(x)dx = \int_0^{k} xf(x)dx + \int_k^{\infty} xf(x)dx \geq \int_k^{\infty} xf(x)dx \geq \int_k^{\infty} kf(x)dx = k\int_0^{\infty} f(x)dx = kp[x \geq k]$$

٤-٩-٢- متباينة Inégalité de Tchebychev- Bienayme

تشبيشيف بينامية :

نفرض أن x متغير عشوائي توقعه μ وانحرافه المعياري Б ، فإنه مهما يكن ε (epsilon) عدد موجب تماما $\varepsilon>٠$ ، يكون

$$p\left[|x-\mu| \geq \varepsilon\right] \leq \frac{\sigma^2}{\varepsilon^2}$$

البرهان: بملاحظة أن $(x-\mu)^2$ متغير عشوائي يأخذ فقط قيم غير سالبة فيمكن تطبيق متباينة ماركوف، بوضع $k=\varepsilon^2$ فنجد :

$$p\left[(x-\mu)^2 \geq \varepsilon^2\right] \leq \frac{E(x-\mu)^2}{\varepsilon^2}$$ بما أن

$$(x-\mu)^2 \geq \varepsilon^2 \Leftrightarrow |x-\mu| \geq \varepsilon$$

مثال : نفرض أن مجتمعا من الأشخاص حيث القامة المتوسطة تساوي ١,٦٥م مع تباين يساوي $\frac{1}{100}$ م، برهن أن الاحتمال أن تنحصر قامة شخص ما بين ١,٤٥ م و ١,٨٥م تساوي على الأقل 75 %؟

حل المثال:

$$|x-\mu| = 0.2, \sigma^2 = 0.01m, \quad \mu = 1.65m,$$
$$p(1.45 < x < 1.85) = p(|x-\mu| < 0.2) = 1 - p\left[|x-\mu| \geq 0.2\right]$$

باستخدام متباينة تشييشييف حيث

$$p\left[|x-\mu| \geq 0.2\right] \leq \frac{\sigma^2}{(0.2)^2} = \frac{0.01}{0.04} = \frac{1}{4}$$

ومنه يكون $p(1.45 < x < 1.85) = 1 - \frac{1}{4} = \frac{3}{4} = 0.75$

تمارين: المتغيرات العشوائية والتوزيعات الاحتمالية:

تمرين رقم ١:

اختيرت عينة من ثلاث وحدات بطريقة عشوائية من صندوق به ١٠ وحدات بينها ٣ وحدات معيبة، أكتب جدول تبين به قيم المتغير العشوائي لعدد الوحدات المعيبة؟

تمرين رقم ٢: اذا كان عمر نوع من المصابيح التي تنتجها مؤسسة" سيقما" له دالة كثافة احتمالي كما يلي:

$$f(x) = \begin{cases} ke^{2x} ; 0 \leq x \leq \ln 2 \\ 0 / w \end{cases}$$

احسب الثابت k ؟

تمرين رقم ٣: في تجربة إلقاء حجري نرد، نفرض أن x و y هما المتغيران العشوائيان المعرفان على الفضاء Ω ، ونفرض أن x هو الذي

يعطي القيمة الكبرى للعددين في أي نقطة من Ω، وأن y هو الذي يعطي جداء العددين في أي نقطة من Ω،

- أوجد توزيع x و y في جدول ؟
- أوجد التوزيع الاحتمالي الهامشي للمتغيرين x و y؟
- خل المتغيرين x و y مستقلة أم لا؟

<u>تمرين رقم ٤:</u> إن دالة الاحتمال المزدوجة لمتغيرين عشوائيين مستمرين x و y هي كما يلي:

$$f(x, y) = \begin{cases} cx^2 y^2 ; 0 < x < 2 ; 2 < y < 3 \\ 0 / w \end{cases}$$

- أحسب الثابت C، ثم أوجد احتمال P (x = 1, y= 2)
- أوجد التوزيع الهامشي لكل من x و y ؟
- أوجد التوزيع الشرطي لكل من x و y ؟
- هل x و y مستقلين أم لا؟

<u>تمرين٥:</u> أحسب التوقع الرياضي والتباين في كل من التمارين الأربعة السابقة.

٥- التوزيعات لاحتمالية الشهيرة Les Distributions usuelles:

٥-١- التوزيعات للاحتمالية المتقطعة Lois Discrétes:

٥-١-١- توزيع برنولي Loi de Bernoulli:

ويعرف هذا التوزيع بتجربة برنولي وتكون نتيجتها إما نجاحا (Succès) أو فشل (échec) ، نرمز للنجاح بـ(P) والفشل (q) حيث p+q=1
التوزيع الاحتمالي للمتغير العشوائي (x) يكتب على النحو التالي:

$$P(X=x)=\begin{cases} p^{x}(1-p)^{1-x}, & x_i=0,1 \\ 0/w \end{cases}$$

أما توقع وتباين هذا التوزيع هو كما يلي:

$$\mu=E(x)=p, \quad \sigma^{2}=v(x)=pq$$

أما الدالة المولدة للعزوم فهي كما يلي:

$$M_{x}(t)=q+pe^{t}$$

<u>ملاحظة:</u> الإثبات يترك للطالب كواجب.

٥-١-٢- التوزيع الثنائي : Loi Binominale:

هو تعميم لتوزيع برنولي عندما تكون $n\geq 1$ ويتميز بالخصائص التالية:

١- التجارب تتألف من عدد من المحاولات، ونفرضه n والذي يمثل ججم العينة.

٢- المحاولات مستقلة عن بعضها.

٣- لكل محاولتين نتيجتين فقط (النجاح/ الفشل).

٤- احتمال انجاح متساوي وثابت لجميع المحاولات.

قانون احتمال x يسمى ثنائي الحد ذا الوسيطين n و P، ونرمز له بـ Bi (n, P) ، أما التوزيع الاحتمالي فيأخذ الشكل التالي:

$$P(X=x)=\begin{cases} C_n^x p^x (1-p)^{1-x}, & x_i = 0,1...n \\ 0/w \end{cases}$$

أما توقع وتباين هذا التوزيع فهو كما يلي:

$$\mu = E(x) = np, \quad \sigma^2 = v(x) = npq$$

أما الدالة المولدة للعزوم فهي كما يلي:

$$M_x(t) = (q + pe^t)^n$$

<u>**ملاحظة:**</u> الإثبات يترك للطالب كواجب.

٥-١-٣- توزيع بواسون Loi de poisson:

نقول عن متغير عشوائي x أنه يتوزع وفق توزيع بواسون بوسيط $\lambda > 0$ ، ونكتب $p(x;\lambda)$ $x \mapsto$ ، إذا كانت له دالة الاحتمال التالية:

$p(x;\lambda) = P(X = x) = e^{-\lambda}\dfrac{\lambda^x}{x!}; \quad x_i = 0,1\ldots$ ، نلاحظ أن :

$$p(x;\lambda) = e^{-\lambda}\frac{\lambda^x}{x!} \geq 0$$

$$\sum_{x=0}^{\infty} p(x;\lambda) = \sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}(e^{\lambda}) = 1$$

لان $e^x = \sum_{n=0}^{\infty}\dfrac{x^n}{n!}$ (سلسلة صحيحة série entier)

أما تباين وتوقع هذا التوزيع فهو كما يلي: E(x)= V(x)= λ

<u>ملاحظة:</u> إثبات التوقع والتباين يجب الإلمام بمعرفة قواعد السلاسل الصحيحة(أنظر أكثر الى مقياس التحليل الرياضي).

أما الدالة المولدة للعزوم بالنسبة لتوزيع بواسون فهي كما يلي:

$$M_x(t) = e^{\lambda(e^t - 1)}; \quad \forall t \in \mathbb{R}$$

ولتوزيع بواسون تطبيقات واسعة، فهو يقدم بشكل عام نموذجا للمعلومات الإحصائية التي تأخذ شكل تعداد لحوادث نادرة الوقوع، حيث يمثل

المتغير العشوائي x عدد الحوادث النادرة الملحوظة في وحدة قياس معينة زمنيا، بينما يمثل λ معدل أو متوسط عدد مرات ظهور تلك الأحداث في وحدة القياس، كذلك يستخدم توزيع بواسون كتقريب للتوزيع الثنائي عندما ما تكون n كبيرة $(n \to \infty)$ واحتمال النجاح ضئيل يقترب من الصفر $(p \to 0)$.

<u>مثال:</u> ليكن متوسط عدد المكالمات الهاتفية التي يتلقاها مستقبل المكالمات في مركز هاتفي معين ما بين الساعة ٩سا و ١٠سا هو ١,٥ مكالمة في الثانية .

<u>**المطلوب :**</u> حساب احتمال أن يكون لدينا ما بين (١٠سا و ٣٣د) و (١٠سا و ٣٤د):

١) عدم وجود أي مكالمة هاتفية؟

٢) مكالمة هاتفية واحدة؟

٣) مكالمتان هاتفيتان ؟

٤) ثلاث مكالمات هاتفية على الأقل؟

حل المثال:

إن عدد المكالمات الهاتفية x هو متغير عشوائي بواسوني وسيطه $\lambda = 1.5$، وتكون له دالة الاحتمال التالية:

$$p(x;1.5) = P(X = x) = e^{-1.5} \frac{(1.5)^x}{x!}; \quad x_i = 0,1...$$

١)احتمال عدم تلقي أي مكالمة:

$$p(0;1.5) = P(X = 0) = e^{-1.5} \frac{(1.5)^0}{0!} = 0,2231$$

٢) احتمال تلقي مكالمة واحدة:

$$p(1;1.5) = P(X = 1) = e^{-1.5} \frac{(1.5)^1}{1!} = 0,3347$$

٣) احتمال تلقي مكالمتين:

$$p(2;1.5) = P(X = 2) = e^{-1.5} \frac{(1.5)^2}{2!} = 0,2510$$

احتمال تلقي ثلاث مكالمات على الأقل:

$$p(x \geq 3) = 1 - p(x < 3) = 1 - [p(x=0) + p(x=1) + p(x=2)] = 1 - [0,2231 + 0,3347 + 0,2510] = 0,1912$$

ملاحظة: نكتفي بدراسة هذه التوزيعات الثلاثة بالنسبة للتوزيعات المتقطعة، وهناك توزيعات أخرى لم ندرسها من بينها التوزيع الهندسي (توزيع باسكال)، التوزيع الهندسي الزائد....

٥-٢- التوزيعات الاحتمالية المستمرة: Distribution probabilité continues:

هناك عدة توزيعات سنكتفي بدراسة ثلاث توزيعات مهمة (التوزيع غاما، التوزيع الأسي، التوزيع الطبيعي).

٥-٢-١ توزيع غاما:

يعد هذا التوزيع الأساس النظري لتوليد بعض التوزيعات الاحتمالية كالتوزيع الأسي مثلا، ويعالج هذا التوزيع عادة المتغيرات العشوائية التي تكون قيمتها موجبة دائما، ومن أمثلة هذا التوزيع فترات الانتظار على سبيل إجراء تجارب الحياة، الفترة الزمنية لبقاء إنسان على قيد الحياة مصاب بمرض مزمن)،......... وبافتراض أن لدينا متغير عشوائي مستمر وليكن x ، يتوزع وفق توزيع غاما، فإن دالة التوزيع الاحتمالي له تأخذ الشكل التالي:

$$f(x) = \begin{cases} \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & , x > 0 \\ 0 / w & \end{cases}$$

مع العلم أن $\alpha, \beta > 0$

حيث أن α و β تمثل معلمات هذا التوزيع.

$\Gamma(\alpha)$ تمثل دالة غاما وتأخذ الشكل التالي:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy$$

وغالبا ما يعبر عن توزيع غاما باختصار بالاصطلاح الآتي

$x \mapsto G(\alpha,\beta)$ ، وهذه الدالة تتمتع بالخصائص التالية:

$$a)\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$$

$$b)\Gamma(1) = 1, \quad c)\Gamma(\frac{1}{2}) = \sqrt{\pi}$$

أما توقع هذا التوزيع وتباينه هو كما يلي:

$$\mu = \alpha\beta, \quad v(x) = \alpha\beta^2$$

<u>مثال:</u> إذا كان لدينا المتغير العشوائي المستمر (x) والذي يمثل الفترة الزمنية لعمل آلة إنتاجية (بالسنين)، نأخذ التوزيع الاحتمالي التالي:

$$f(x) = \begin{cases} \frac{x}{16}e^{-\frac{x}{4}} & x \geq 0 \\ 0 \,/\, w & \end{cases}$$

<u>المطلوب:</u>

١) إثبات دالة التوزيع الاحتمالي لهذا المتغير هي دالة كثافة احتمالية؟

٢) ما هو احتمال أن تستمر الآلة بالعمل لمدة (١٠ سنوات) أخرى على الأكثر؟

٣) حساب متوسط عمر الآلة (μ_x) وكذا التباين $v(x)$؟

حل المثال:

١)لإثبات أن دالة $f(x)$ هي دالة كثافة احتمالية ينبغي أن تحقق الالة الخاصيتين الآتيتين:

$$a)\ f(x) \geq 0, \quad [\forall x, x \geq 0]$$

$$b)\int_0^{\infty} f(x)dx = 1$$

الحالة الأولى محققة $a)\int_0^{\infty} \frac{x}{16} e^{-\frac{x}{4}} dx$

الحالة الثانية نضع $b)\ y = \frac{x}{4}, \quad dy = \frac{dx}{4}$

نستخدم التكامل بالتجزئة ونتحصل على١:

$$\frac{1}{16}\int_0^{\infty} 4ye^{-y}4dy = \int_0^{\infty} ye^{-y}dy$$

$$2)p(x \leq 10) = -[0,000045-1] = 0,999$$

$$3)\mu_x = \alpha\beta = (2)(4) = 8ans, \quad v(x) = \alpha\beta^2 = 2(16) = 32 \quad ans$$

٥-٢-٢- التوزيع الأسي: La loi exponentielle

بعد هذا التوزيع حالة خاصة من توزيع غاما $(\alpha = 1)$، ويستخدم في معالجة بعض التطبيقات الإحصائية مثل تقدير مدة حياة بعض الأجهزة، فترات الانتظار، درجات الحرارة العظمى والصغرى...وبافتراض أن لدينا متغير عشوائي مستمر (x) يتوزع وفقا للتوزيع الأسي، فإن دالة التوزيع الاحتمالي تأخذ الشكل الآتي:

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

وبوضع $\lambda = \frac{1}{\beta}$، يصبح لدينا:

$$f(x) = \begin{cases} \lambda e^{-\lambda} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

أما خواص هذا التوزيع فنوجزها في هذه النقاط:

$$1) p(x > a) = e^{-\lambda a}, \quad 2) p(x < a) = 1 - e^{-\lambda a}, \quad 3) p(a < x < b) = e^{-\lambda a} - e^{-\lambda b}$$

<u>مثال:</u>

نفرض أن زمن مكالمة هاتفية مقاسة بالدقائق هو متغير عشوائي أسي وسيطة $\lambda = \frac{1}{20}$، يصل الشخص A إلى حجرة الهاتف، وفي نفس اللحظة يمر قبله شخص آخر (دخل إلى الحجرة).

١) ما هو احتمال أن الشخص A ينتظر أكثر من ٢٠ دقيقة؟

٢) ما هو احتمال أن الشخص A ينتظر ما بين ٢٠ و ٤٠ دقيقة؟

حل المثال:

ليكن X المتغير العشوائي الذي يمثل زمن المكالمة الهاتفية X هو متغير أسي وسيطه $\lambda = \frac{1}{20}$.

١) الحادث "الانتظار الأكثر من ٢٠ دقيقة هو (X>20) واحتماله هو:

$$1)\ p(x > 20) = e^{-\frac{1}{20}(20)} = e^{-1} = \frac{1}{e}$$

٢) الحادث "الانتظار ما بين ٢٠ و ٤٠ دقيقة "هو $(20 < x < 40)$ والذي احتماله هو

$$2)\ p(20 < x < 40) = e^{-\frac{20}{20}} - e^{-\frac{40}{20}} = e^{-1} - e^{-2}$$

أما التوقع الرياضي والتباين للمتغير العشوائي الأسي هو:

$$E(x) = \frac{1}{\lambda}, \quad v(x) = \frac{1}{\lambda^{2}}$$

٥-٢-٣- التوزيع الطبيعي La loi normale:

يعد التوزيع من أهم التوزيعات الاحتمالية المستمرة، إذ يستخدم على نطاق واسع في وصف عدد كبير من الظواهر الطبيعية والاجتماعية والاقتصادية......

تعريف: نقول عن متغير عشوائي x أنه يخضع للتوزيع الطبيعي إذا كانت كثافة توزيعه الاحتمالي معرفة كما يلي:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \qquad -\infty < x < +\infty$$

σ : وسيط يساوي الانحراف المعياري للتوزيع.

μ : وسيط يساوي الانحراف التوقع الرياضي للتوزيع.

ونرمز له باختصار x ~ N (σ^2 μ,)

٥-٢-٣-١- التوزيع الطبيعي المعياري:

إذا كان المتغير العشوائي المستم (x) يتوزع وفق التوزيع الطبيعي بوسط حسابي (μ) وتباين (σ^2) يمكن الحصول عليه من خلال إجراء التحويل

الآتي: $Z=\dfrac{x-\mu}{\sigma}$، وعليه فدالة الكثافة الاحتمالية تعطى بالشكل الآتي:

$$\Phi(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^{2}} & -\infty < x < +\infty \\ 0 / w & \end{cases}$$

أي أن التوزيع الطبيعي المعياري هو حالة خاصة من التوزيع الطبيعي ويكون فيها $\mu=0$ و $\sigma^{2}=1$.

٥-٢-٣-٢- دالة التوزيع الطبيعي المعياري:

بشكل خاص نرمز بـ $(\Phi(x))$ لدالة التوزيع المتغير العشوائي x الذي يخضع للتوزيع الطبيعي المعياري أي أن :

$$\Phi(z)=\int_{-\infty}^{z} e^{-\frac{1}{2}x^{2}}\,dx; \quad z \in \mathbb{R}$$

إن هذا التكامل يحسب بشكل تقريبي، ولذلك فقد أعد جدول يعطي قيم $\Phi(z)$ والتي تمثل المساحة المحددة.

<u>مثال:</u> إذا علمت أن أطوال مجموعة من الطلبة تتوزع طبيعيا بالوسط ١٦٠ سم وبانحراف معياري ٥ سم، أوجد احتمال اختيار طالب يكون طوله ما بين ١٥٥سم و ١٦٥سم.

حل المثال:

لحل هذا المثال علينا إجراء عملية التحويل التوزيع الطبيعي إلى التوزيع الطبيعي المعياري باستخدام صيغة القيمة المعيارية السابقة $\sigma = 5cm$ ، $\mu = 160cm$ ، $Z = \frac{x-\mu}{\sigma}$ وبعد ذلك يتم حساب الاحتمال وهو كما يلي:

$$p(155 < x < 165) = p(\frac{150-160}{5} < Z < \frac{165-160}{5}) = p(-1 < z < 1) = 0,3413 * 2 = 0,6828$$

أما الدالة المولدة للعزوم بالنسبة للتوزيع الطبيعي المعياري فهي كما يلي:

$$M_x(x) = e^{\frac{t^2}{2}}$$

ملاحظة: يمكن إجراء عملية التقريب التوزيع الثنائي بالتوزيع الطبيعي بواسطة العلاقة التالية: $Z = \frac{x - np}{\sqrt{npq}}$

تمارين: التوزيعات الاحتمالية الشهيرة:

تمرين١ : احتمال أن يكسب الفريق A في أي مباراة يلعبها هو $\frac{2}{3}$، إذا لعب هذا الفريق أربع مباريات، فأوجد احتمال أن يكسب:

١) مبارتين بالضبط ، ٢) أكثر من نصف المباريات ، ٣) على الأقل مباراة واحدة؟

تمرين٢ : نفرض أن ٣% من مواد التنظيف التي ينتجها مصنع ما، غير صالحة، سحبت عينة بالصدفة تتألف من ٢٠٠ مادة تنظيف من المواد التي ينتجها هذا المصنع، أحسب احتمال أن نجد في العينة ٤ مواد غير صالحة؟

تمرين٣: التوزيع التالي يطلق عليه التوزيع الطبيعي المعياري :

$$\Phi(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} & -\infty < x < +\infty \\ 0 / w & \end{cases}$$

أوجد الدالة المولدة للعزوم لهذا التوزيع؟ أحسب التوقع والتباين؟

تمرين ٤ :

أ) _نفرض أن X متغير عشوائي وأن له التوزيع الطبيعي المعياري Φ عين قيم m في كل حالة:

1) $p(0 \leq x \leq m) = 0.4656$

2) $p(x \leq m) = 0.9750$

3) $p(m \leq x \leq 1.5) = 0.7745$

4) $p(x \leq m) = 01469$

ب) أ) نفرض أن أنصاف البراغي التي ينتجها أحد المصانع موزعة توزيعا طبيعيا بتوقع 25mm و بانحراف معياري 2mm ، يعتبر البرغي معيبا إذا كان نصف قطره يقل عن 20mm أو يكبر عن28mm ، أوجد احتمال أن يكون البرغي معيبا؟

٦- طرق المعاينة وتوزيعاتها:

تمهيد: ان الهدف من الاحصاء الوصفي هو تحويل عدة ارقام ومعطيات الى شكل موجز وواضح يعبر عن اهم ما يوجد في هذه المعطيات، أما الاحصاء الاستدلالي فالغاية منه هو تحليل المشاهدات الاحصائية لاستخلاص بعض المعلومات عن مجتمع، والذي هو عبارة عن جميع المفردات التي يمكن أن يأخذها المتغير، أما العينة فهي جزء من المجتمع لان دراسة هذا الأخير تكون في بعض الاحيان مستحيلة او مكلفة وتحتاج الى جهد ووقت طويل.

نريد دراسة الادخار في الجزائر يمكننا أخذ عينة من أرباب الأسر، وهذه العينة ستعطينا " فكرة " عن الادخار الوطني، والمشكل هو البحث عن أحسن وسيلة لمعرفة هذا الادخار.

<u>تعريف:</u>

اذا كانت $x_1, \quad x_2 ...... \quad x_n$ ،عينة عشوائية من الحجم n للمتغير العشوائي x ، عندئذ متوسط العينة بالتعريف هو الاحصاءة : $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$، أما التباين العينة بالتعريف هو الاحصاءة : $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$

٦-١- توزيعات بعض الاحصاءات:

٦-١-١- توزيعات المعاينة للمتوسطات:

مثال١: نفرض أن مجتمع ما يتكون من ١٠٠٠ عنصر، بوسط حسابي ٣٠ وحدة و انحراف معياري ١٢ وحدة، المطلوب: إيجاد توزيع المعاينة لوسط عينة حجمها ٤٩.

حل مثال١: $\mu_{\bar{x}} = \mu = 30$ ، $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{49}} = 1.71$

ملاحظة: إذا كانت **(n>0.05N)** ندخل ما يعرف بمعامل التصحيح، في مثالنا السابق لو كانت n تساوي ٨١ بدلا من ٤٩ فأن

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\bar{x}} = \frac{12}{\sqrt{81}} \sqrt{\frac{1000-81}{1000-1}} = 1.27$$

مثال٢: يمكن حساب احتمال أن يقع وسط عينة عشوائية $\bar{x}$ حجمها ٤٩ بين ٢٨ و ٣٤ (خاصة بالمعطيات المثال السابق).

حل مثال٢:

$$p(28<\bar{x}<34)=p(\frac{28-30}{1.71}<Z<\frac{34-30}{1.71})=p(-1.71<z<2.34)=0.3799+0.4904=0.8703$$

٦-١-٢- توزيع المعاينة للفرق بين وسطين

أ) سحبت عينة عشوائية حجمها n_1 من مجتمع طبيعي وسطه (μ_1) وتباينه (σ_1^2) وعينة ثانية من مجتمع طبيعي أيضا وسطه (μ_2) وتباينه (σ_2^2)، والعينة الثانية مستقلة عن الاولى ، فإذا كان ($\bar{x}_1$) يمثل الوسط الحسابي للعينة الاولى و ($\bar{x}_2$) يمثل الوسط الحسابي للعينة الثانية، فإن توزيع المعاينة للفرق بين متوسطيهما $(\bar{x}_1-\bar{x}_2)$ يتبع التوزيع الطبيعي ذا الوسط $(\mu_1-\mu_2)$ و التباين $\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}$ وأن توزيع المعاينة يكون $(\bar{x}_1-\bar{x}_2)\rightarrow N(\mu_1-\mu_2,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$، هذه النتيجة تكون بالنسبة لمجتمعات منتهية أو إذا كان السحب بدون إعادة ، في هذه الحالة القيمة (المتغير) المعيارية تكون كالتالي

$$Z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

ب) حالة مجتمعين غير طبيعيين وحجم العينة كبيرين $(n_1, n_2 \geq 30)$

تكون القيمة المعيارية كالتالي $Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ ، نستبدل

تباين المجتمعين بتباين العينتين S_1^2 ، S_2^2

مثال: سحبت عينتين عشوائيتين من مؤسستين تتبع التوزيع الطبيعي وكان متوسط الأجر المدفوع من قبل المؤسسة (أ) ل ٤٩ عاملا يساوي ٢٣٠٠٠ دج بانحراف معياري ٣٢٠٠ دج ، أما متوسط الأجر المدفوع من قبل المؤسسة (ب) ل ٦٤ عاملا يساوي ٢١٠٠٠ دج بانحراف معياري ١٥٠٠ دج، فأحسب احتمال أن متوسط الأجر المدفوع من قبل المؤسسة (أ) لها متوسط على الأقل ٣٠٠٠ فأكثر من متوسط الأجر المدفوع من قبل المؤسسة (ب)

حل المثال:

$n_1 = 49, \mu_1 = 23000DA \quad , \quad \sigma_1 = 3200DA$

$n_2 = 64, \mu_2 = 21000DA \quad , \quad \sigma_1 = 1500DA$

$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2 = 23000 - 21000 = 2000DA$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(3200)^2}{49} + \frac{(1500)^2}{64}} = 494,10$$

$$P(\bar{x}_1 - \bar{x}_2 \geq 6000) = p(z \geq \frac{3000 - 2000}{494,10}) = p(z \geq 2) = 0,5 - 0,4772 = 0,0228$$

ج) حالة العينتين صغيرتين $(n_1 \, ou \quad n_2 < 30)$ والمجتمعين اللذان سحبتا منه العينتين طبيعيين، تكون القيمة المعيارية كالتالي:

$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SP\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ ، نستخدم توزيع t بدرجة حرية (n_1+n_2-2) ,

Sp هو الانحراف المعياري المرجح (pondéré) حيث

$$SP^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

<u>٦-٢-١- توزيع المعاينة للنسب</u>

نفرض أن مجتمعا ما لا نهائي، ويتوزع وفقا للتوزيع الثنائي حيث أن احتمال النجاح p واحتمال الفشل $q=1-p$ ، وعند اعادة التجربة n من المحاولات وبذلك فان توزيع المعاينة للمتغير العشوائي x المتمثل بعدد النجاحات في العينات ذات حجم n يمكن أن يكون قريبا من التوزيع الطبيعي بوسط حسابي قدره $\mu = np$ و بانحراف معياري $\sigma = \sqrt{npq}$ ،

علما أ، نسبة النجاح مختلفة من عينة الى أخرى ، يمكن التعبير عن نسبة النجاح p بالمقدر $\hat{p}=\frac{x}{n}$

حيث أن x : عدد محاولات النجاحات ، n: حجم العينة ، $\hat{p}$: نسبة النجاح في العينة.

فتوزيع المعاينة ل $(\hat{p})$نسبة النجاحات هو قريب من التوزيع الطبيعي بوسط $\mu_{\hat{p}}=np$ ،

وانحراف معياري $\sigma_{\hat{p}}=\sqrt{\frac{pq}{n}}$، والقيمة المعيارية Z هي $Z=\frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$

مثال:

إذا كان إحتمال نجاح طالب في مسابقة ما هو ٠,٦ ، سحبت عينة عشوائية حجمها ٦٤ طالبا، أوجد احتمال أن يكون $p(0,5<\hat{p}<0,7)$ ؟

حل المثال: n=64، p=0.6

$$p(0,5<\hat{p}<0,7)=p(\frac{0,5-0,6}{\sqrt{\frac{(0,6)(0,4)}{64}}}<Z<\frac{0,7-0,6}{\sqrt{\frac{(0,6)(0,4)}{64}}})=p(-1,63<Z<1,63)=0,4484*2=0,8968$$

٦-٢-٢- توزيع المعاينة للفرق بين نسبتين:

سحبت عينتان عشوائيتان حجمهما n1 ، n2 من مجتمعين مستقلين، يخضع الأول والثاني للتوزيع $B_i(n_1, p_1)$ و $B_i(n_2, p_2)$ وأن

$\sigma_1^2 = n_1 p_1 q_1$ و $\mu_1 = n_1 p_1$

و $\sigma_2^2 = n_2 p_2 q_2$ و $\sigma_1^2 = n_1 p_1 q_1$ ، فتوزيع المعاينة للفرق بين $(\hat{p}_1 - \hat{p}_2)$ يقترب من التوزيع الطبيعي بوسط $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$ و

بانحراف معياري $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$ والقيمة المعيارية لهما

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

<u>مثال:</u> إذا كانت نسبة النجاح في مقياس الاحصاء في كلية العلوم الاجتماعية في جامعة (أ) تساوي ٠,٩ وكانت نسبة النجاح انفس المقياس من نفس الكلية في جامعة (ب) تساوي ٠,٨، سحبت عينة عشوائية حجمها ١٤٠ طالب من جامعة (أ) و عينة ثانية عشوائية حجمها ٨٠ طالب من جامعة (ب)، أوجد إحتمال ان تزيد نسبة النجاح في جامعة (أ) عن

نسبة النجاح في جامعة (ب) بمقدار ٠,١٥ على الأقل.

حل المثال:

$$n_1 = 140, p_1 = 0,9 \quad , \quad q_1 = 0,1$$
$$n_1 = 80, p_2 = 0,8 \quad , \quad q_2 = 0,2$$

المطلوب هو $p(\hat{p}_1 - \hat{p}_2 \geq 0,15)$

$$p(\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \geq \frac{0,15 - (0,9 - 0,8)}{\sqrt{\frac{(0,9)(0,1)}{140} + \frac{(0,8)(0,2)}{80}}}) = p(Z \geq 0,97) = 0,5 - 0,3340 = 0,166$$

٧- التقديرات Les estimations

تمهيد: الهدف الأساسي للإحصاء الاستدلالي هو الاستقراء ، أو التنبؤ بمعلومات من المجتمع من واقع معلومات محتواه في العينة ، والخطوة الأساسية هي وصف مجموعة من المشاهدات مأخوذة من المجتمع سواء عن طريق تمثيلها بالرسم البياني أو بالتوزيعات التكرارية، والخطوة الهامة هي استخدام بيانات العينة للاستقراء أو التنبؤ بخصائص المجتمع محل الدراسة، لذلك لابد من معرفة التوزيعات الاحتمالية المتعلقة بالعينة

والمقاييس الاحصائية الاخرى لتساعدنا في تحليل البيانات للوصول الى استقراء أو تنبؤ جيد، وتعد التقديرات من الأدوات المهمة في هذه المرحلة، فربما نود تقدير متوسط دخل الأسرة في مجتمع أو نسبة الناخبين الذين يفضلون مرشح معين... لذلك لابد من فهم جيد لأساليب اختيار التقديرات وخصائصها ومدى ملائمتها لتقدير معالم المجتمع، وهناك نوعان من التقديرات :

أ) التقدير النقطي L'estimation ponctuelle

ب) التقدير بمجال ثقة L'estimation par L'intervalle de la confiance

<u>٧-١- التقدير النقطي:</u>

هو القيمة العددية لاحصاءة العينة الذي يستخدم لتقدير القيمة لمعلمة المجتمع، لنفرض أن قانون توزيع مجتمع ما يدل عليه بمعلمة θ تنتمي إلى فضاء Ω فالمسألة هي تقدير θ على أساس عينة من المجتمع.

مثال١: إذا كان توزيع المجتمع طبيعيا $N(\mu,\sigma^2)$ وإذا μ و σ^2 مجهولان ، فان $\theta=(\mu,\sigma^2)$ ، فان $\theta=\mu$ و $\Omega=]-\infty,+\infty[\text{X}]0,+\infty[$ إن المقدر

(estimateur) يدل على إحصائية قيمتها هي التقدير للمعلمة θ، فمثلا $\bar{x}$ هو مقدر لمتوسط المجتمع وهكذا.

مثال٢: $\bar{x}$ هو مقدر غير متحيز (non bias) للوسط μ مهما كان التوزيع، أما $\hat{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2}{n}$ ، فأنه مقدر متحيز للتباين σ^2.

<u>٧-٢- التقدير بمجال ثقة:</u>

بمأن التقديرات النقطية تعتمد على قيم العينة فقط، فهي لا تنطبق على معالم المجتمع(لأنها مؤشر لأحدى المقادير المجهولة)، وسيبق هذا المؤشر عديم الفائدة مادام غير مصحوب بمجال الثقة، فاذا قلنا مثلا في دراسة لعينة ما عدد مفرداتها n ، ان متوسط هذه العينة هو ٢٥ وحدة، فهذا الرقم لا يدل بالضرورة على أنه هو المتوسط الحسابي لكل المجتمع الذي أخذت منه العينة ، ولكن يمكننا إعطاء فكرة وهي أنه في الامكان أن يكون المتوسط الحسابي للمجتمع ٢٥ أو قريب منه، وكما يجعلنا نفكر أن المتوسط الحسابي للمجتمع يتراوح بين هذه القيمة بالزيادة أو النقصان، وبعبارة أدق إن القيمة التي تحصلنا عليها من قياس العينة تسمح لنا بالقول أن هناك نسبة مئوية ولتكن مثلا ٩٩%من الحظ على أن المتوسط

الحسابي للمجتمع يتراوح بين القيمة ٢٥ وقيمة أخرى نحددها مسبقا، وهو ما يعرف بالمجال.

٧-٣- تقدير مجال الثقة للمتوسط:

٧-٣-١- مجال الثقة لمتوسط متغير عشوائي طبيعي تباينه معلوم:

ليكن x متغيرا عشوائيا طبيعيا متوسطه μ مجهول، وتباينه σ^2 معلوم، ولتكن $x_1, x_2, \ldots\ldots\ldots, x_n$، عينة عشوائية لـx ، المطلوب إيجاد مجال ثقة للعلمة μ، يمكن إيجاد توزيع المعاينة

$\bar{x} \rightarrow N(\mu \quad , \quad \frac{\sigma^2}{n}) \Rightarrow \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \rightarrow N(0, \quad 1)$، حيث يخضع للتوزيع الطبيعي المعياري N(0,1) b,a بحيث (a < b) بحيث :

$P\left[a \leq N(0, \quad 1) \leq b\right] = 1-\alpha \ldots\ldots\ldots(1)$ نستطيع كتابة (١) بالشكل التالي:

$$P_{\mu}\left[a \leq \frac{\sqrt{n}\,(\bar{x}-\mu)}{\sigma} \leq b\right] = 1-\alpha$$ العبارة الأخيرة تكافئ:

$$P_{\mu}\left[\bar{x}-b\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x}-a\frac{\sigma}{\sqrt{n}}\right]$$ ، ونتيجة لذلك(٢)

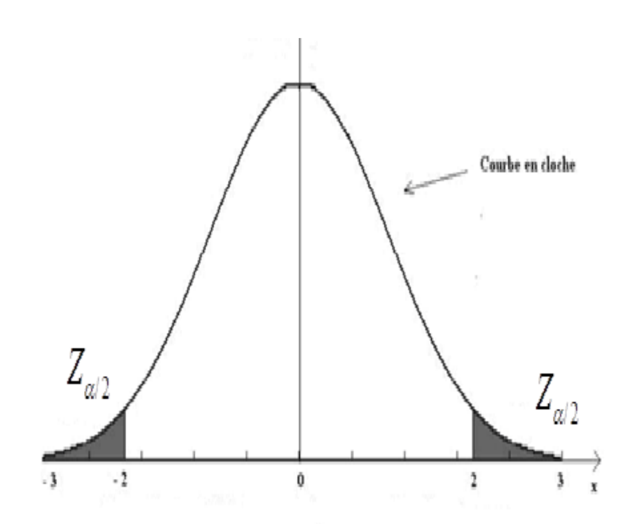
$$\Delta_y(x) = \left[\bar{x}-b\frac{\sigma}{\sqrt{n}} ; \bar{x}-a\frac{\sigma}{\sqrt{n}}\right] \ldots\ldots\ldots$$

هو مجال ثقة للمتوسط μ من الدرجة % $(1-\alpha)$ 100 أي نقول أننا واثقون بمقدار $1-\alpha$ من أن μ لن يقل عن $\bar{x} - b\frac{\sigma}{\sqrt{n}}$ ولن يزيد على $\bar{x} - a\frac{\sigma}{\sqrt{n}}$.

وهكذا فإن العبارة ٢ يمكن كتابتها على النحو التالي:

$$\left[\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

والشكل البياني التالي يوضح ذلك:



والجدول التالي يقدم الأربع معاملات الثقة الأكثر شيوعا لقيم

$1-\alpha$	α	$\alpha/2$	$Z_{\alpha/2}$
٠,٩٠	٠,١	٠,٠٥	$Z_{0.01}$= 1.64
٠,٩٥	٠,٠٥	٠,٠٢٥	$Z_{0.025}$= 1.96

$Z_{0.01}= 2.33$	٠,٠١	٠,٠٢	٠,٩٨
$Z_{0.005}= 2.57$	٠,٠٠٥	٠,٠١	٠,٩٩

مثال: أجريت إحدى شركات الهاتف النقال دراسة حول مدة استخدام الهاتف، لذا سحبت عينة عشوائية مكونة من ١٠٠٠٠ نداء ووجد أن مدة استخدامهم الهاتف كانت ٤ دقائق، أما الانحراف المعياري للمجتمع بلغ ١,٥ دقيقة، قدر قيمة الوسط الحسابي للمجتمع بحدود ثقة ٩٥%.

حل المثال: $\overline{x} = 4\min, \quad \sigma = 1.5\min, \quad n = 10000$ وباستخدام احتمال مجال ثقة ٩٥% نجد $1-\alpha= 0.95$ ويكون مجال الثقة التقريبي لمتوسط المكالمات μ هو $\left[4-1.96\frac{1.5}{\sqrt{10000}}, 4+1.96\frac{1.5}{\sqrt{10000}}\right]$ أي المجال $[3.985 \quad , \quad 4.015]$

٧-٣-٢- تقدير حجم العينة:

إن اختيار حجم العينة (n) له أهمية كبيرة في النتائج النهائية، فكلما كبرت العينة كانت التقديرات أكثر دقة وأقرب لمعلمات المجتمع. إن المقدار $|\overline{x}-\mu|$ يمثل الخطأ المطلق في التقدير، وتدل العبارة الاحتمالية في هذا التقدير وباحتمال $1-\alpha$ لا يتجاوز المقدار $\pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ زيادة

ونقصانا، فإذا رمزنا بـE للخطأ المطلق المرتكب في التقدير النقطي لـ μ عند مستوى الثقة $1-\alpha$ كان $E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$.

نلاحظ أن الخطأ المطلق يتناقص بازدياد حجم العينة n، ويمكن التحكم به، فإذا أردنا تعيين حجم العينة التي ينبغي اختيارها أخذين بعين الاعتبار أن لا يتجاوز هذا الخطأ $\overline{x} - \mu$ المقدار ε بمجال ثقة $100(1-\alpha)$ % ، فيتم ذلك بحل المتباينة التالية:

$e = \left| \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right| \leq E$ ، وبتربيع الطرفين نحصل على $n \geq (\frac{Z_{\alpha/2}\sigma}{E})^2$

<u>**مثال:**</u> أخذت عينة عشوائية مكونة من ٤٠٠ طالب بكلية العلوم الاجتماعية، ووجد أن متوسط الأوزان كان يساوي ٧٥كلغ، أما الانحراف المعياري للأوزان بالنسبة لجميع طلبة الكلية كان يساوي ٣كغ.

المطلوب: ١- أوجد مجال ثقة ٩٥% لمتوسط المجتمع μ .

٢- كم ينبغي أن يكون حجم العينة بحيث لا يتجاوز الخطأ في التقدير μ وبفترة ثقة ٩٥% المقدار E=0.26.

<u>**حل المثال:**</u> $Z_{\alpha/2} = 1.96 \quad , 1-\alpha = 0.95$

إذن مجال ثقة ٩٥% لمتوسط المجتمع μ هو

$\left[75-1.96\frac{3}{\sqrt{400}};75+1.96\frac{3}{\sqrt{400}}\right]$، نتحصل على المجال

.[74, 706, 75, 296]

٢) تعيين حجم العينة: $n \geq (\frac{Z_{\alpha/2}\sigma}{E})^2 = (\frac{1.96*3}{0.26})^2 = 511.45$

ومنه يجب أن يكون حجم اللازم للعينة $n \geq 512$

٧-٣-٣- حالة التباين σ^2 غير معلوم:

إن معرفتنا لتباين المجتمع σ^2 تبدو ضرورية لتعيين مجال ثقة لـ μ

وعندما يكون Б^2 مجهولا، فإن استبدال تباين المجتمع σ^2 بتباين العينة S^2 يستدعي تفسيرا الحالتين التاليتين:

الحالة الأولى: إن الإحصاءة $\frac{\bar{x}-\mu}{s/\sqrt{n}}$ تحتوي على متغيرين عشوائيين $\bar{x}$ و S، عندما يكون حجم العينة كبيرا $n \geq 30$ فإن تغيرات S^2 من عينة إلى أخرى تكاد تكون معدومة، لذلك يمكن استبدال σ بـS (الانحراف المعياري للمجتمع والعينة على الترتيب).

الحالة الثانية: عندما بكون حجم العينة n < 30 فإن تغيرات S^2 تكون مؤثرة في شكل توزيع الإحصاء $\frac{\bar{x}-\mu}{s/\sqrt{n}}$، ونعرف المتغير العشوائي على

النحو التالي: $t = \frac{\overline{x} - \mu}{s / \sqrt{n}}$ ، التوزيع t (ستيودنت) بـ n−1 درجة حرية.

فرضا أن $(1-\alpha)\%$ 100 مجال الثقة المتوسط μ، إذن نستخدم مجال الثقة كما ورد في الفقرات السابقة، ويكون على النحو التالي:

$$P\left[\overline{x} - \frac{t_{(n-1),\alpha/2} S}{\sqrt{n}} < \mu < \overline{x} + \frac{t_{(n-1),\alpha/2} S}{\sqrt{n}}\right]$$

<u>مثال:</u> أردنا معرفة متوسط استهلاك السيارات من الوقود في السنة مقدرة بـ(ميل/قالون) miles per gallon، فاخترنا (٦) سيارات وكان متوسط الاستهلاك على النحو التالي ٢٠,٨، ١٩,٦، ٢٠,١، ١٨,٥، ١٨,٦ الوقود في المجتمع الذي اخترنا منه السيارات الستة، مفترضا أن استهلاك الوقود في المجتمع بخضع للتوزيع الطبيعي.

<u>حل المثال:</u> يجب حساب المقادير التالية:

$$\overline{x} = \frac{\sum x_i}{n} = \frac{116.6}{6} = 19.43, \quad S = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = 0.6$$

$$t(\alpha/2) = t(\alpha/2, \quad n-1) = t(0.05, \quad 5) = 2.015$$

و باعتبار أن المجتمع يخضع للتوزيع الطبيعي يمكن تشكيل مجال الثقة التالي:

$$P\left[\bar{x}-\frac{t_{\alpha/2}S}{\sqrt{n}}<\mu<\bar{x}+\frac{t_{\alpha/2}S}{\sqrt{n}}\right]=1-\alpha$$

$$P(19.43-(2.015)\frac{0.9}{\sqrt{5}}<\mu<19.43-(2.015)\frac{0.9}{\sqrt{5}})=0.90$$

$$P(18.61<\mu<20.24)=0.90$$

٧-٣-٤- مجال الثقة للفرق بين متوسطي متغيرين عشوائيين:

يمكن استخدام الأسلوب السابق لإيجاد الثقة للفرق بين متوسطين $(\mu_1-\mu_2)$ من مجتمعين مختلفين، باستخدام نظريات المعاينة للفرق بين المتوسطين وفي الحالتين التاليتين:

الحالة الأولى: مجال ثقة للفرق بين متوسطين متغيرين عشوائيين طبيعيين تباينهما معلومان:

فإن توزيع المعاينة لـ $(\bar{x}_1-\bar{x}_2)$ يقترب من التوزيع الطبيعي بوسط:

$$\mu_{\bar{x}_1-\bar{x}_2}=\mu_1-\mu_2$$

و بانحراف معياري $\sigma_{\bar{x}_1-\bar{x}_2}=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

ويكون المتغير العشوائي Z كالتالي: $Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

فإن فترة الثقة ($1-\alpha$) للفرق بين المتوسطين هو كالتالي: نفس الأسلوب المستخدم في الفقرات السابقة:

$$P\left[(\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right] = 1 - \alpha .....(I)$$

<u>ملاحظة:</u> في حالة تباين المجتمعين (σ_1^2 , σ_2^2)غير معلومين n_1 ,n_2 أكبر أو يساوي ٣٠، يمكننا أن نستبدل تباين المجتمعين بتباين العينيتين S_1^2، S_2^2 ، وتبقى فترة الثقة نفسها في (I).

$$P\left[(\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right] = 1 - \alpha$$

الحالة الثانية مجال ثقة للفرق بين متوسطين متغيرين طبيعيين تباينها غير معلوم.

عندما يكون n_1 ,n_2 أقل من ٣٠ ففي هذه الحالة نميز حالتين آخرتين وهما.

أ) **<u>حالة تساوي تباينات المجتمعين</u>** $\sigma_1^2 = \sigma_2^2 = \sigma^2$:

يكون مجال ثقة % $100(1-\alpha)$ للفرق بين المتوسطين $\mu_1-\mu_2$ عندما يكون σ_1^2 و σ_2^2 مجهولين، ولكن شرط أنهما متساويان فإن مجال الثقة:

$$P\left[(\bar{x}_1-\bar{x}_2)-t_{\alpha/2}SP\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+t_{\alpha/2}SP\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]=1-\alpha$$

حيث $\bar{x}_1$ و $\bar{x}_2$ هما متوسط العينيتين العشوائيتين المستقلتين ذات حجم μ_1, μ_2 من مجتمع يقترب من التوزيع الطبيعي و SP هو الانحراف المرجح وأن $t_{(\alpha/2)}$ هي قيمة التوزيع t بدرجة حرية $V= n_1+ n_2- 2$

حيث $SP^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

مثال:

سحبت عينة عشوائية من إحدى المدارس لقياس مستوى التحصيل العلمي ولنفس المستوى، المدرسة الأولى سحبت منها عينة تقدر بـ١٢ تلميذا ثم تدريسهم بالطريقة العادية، والمدرسة الثانية سحبت منها عينة تقدر بـ٨ تلاميذ ثم تدريسهم استخدام جهاز الأيباد (Ipad). وفي نهاية الفصل أعطى نفس الامتحان لكلا المدرستين فكانت النتائج كما يلي:

$n_1= 12, \quad \bar{x}_1 = 14$ ، $S_1^2=$ ٥

$n_2 = 8, \quad \bar{x}_2 = 13.5$ ، $S_2^2= 3.5$

فإذا علمنا أن معدلات التحصيل تتبع التوزيع الطبيعي، وأن لهما نفس التباين أوجد ٩٥% مجال ثقة للفرق بين $\mu_1 - \mu_2$ حيث أن μ_1 معدل التحصيل في المدرسة الأولى و μ_2 معدل التحصيل في المدرسة الثانية، هل هناك فرق حقيقي بين المتوسطين؟

<u>حل المثال:</u>

من جدول توزيع t نجد

$$1-\alpha = 0,95, \Rightarrow \quad \alpha/2 = 0,025, \quad t_{\alpha/2} = t(\alpha/2 \quad , \quad n_1+n_2-2) = t(0.025,18) = 2,101$$

$$SP = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} = \sqrt{\frac{(11)5+(7)3,5}{18}} = 2,101$$

مجال الثقة يكون كالتالي:

$$P\left[(14-13,5)-(2,101)(2,101)\sqrt{\frac{1}{12}+\frac{1}{8}} < \mu_1 - \mu_2 < (14-13,5)+(2,101)(2,101)\sqrt{\frac{1}{12}+\frac{1}{8}}\right] = 0,95$$

أي أن المجال هو [2.51 و 1.5−].

وبملاحظة أن طرفي مجال الثقة مختلفان في الإشارة، فهذه النتائج لا تقدم دلالة على وجود فرق حقيقي بين معدل التحصيل في كلا المدرستين عند مستوى ثقة ٩٥%.

ب) **حالة عدم تساوي تباين المجتمعين** $\sigma_1^2 \neq \sigma_2^2$

في حالات عديدة (عندما يكون تباين المجتمعين غير معلومين مع عدم تساوي تباين هذين المجتمعين $\sigma_1^2 \neq \sigma_2^2$ نستخدم الإحصاءة التالية:

$$T^* = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

فمجال الثقة يكون كالتالي:

$$P\left[(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,v}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,v}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right] = 1 - \alpha$$

t^* يتوزع تقريبا بتوزيع t مع درجات حرية مقدمة بالصيغة التالية:

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}}$$

<u>٧-٤- مجال الثقة للنسب في المجتمع:</u>

كما أشرنا في فقرة سابقة فيمكن التعبير عن نسبة النجاح P بالمقدر

$\hat{p}=\frac{x}{n}$ حيث x عدد المحاولات النجاحات) و n : حجم للعينة، $\hat{p}$ هو

نسبة النجاح في العينة، فتوزيع المعاينة لـ($\hat{p}$) نسبة النجاحات هو قريب

من التوزيع الطبيعي بوسط $\mu_{\hat{p}}=np$ وبانحراف معياري $\sigma_{\hat{p}}=\sqrt{\frac{pq}{n}}$

والقيمة المعيارية Z هي $Z=\frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$ وبنفس استخدام الخطوات السابقة

نجد % $(1-\alpha)$ 100 مجال الثقة لنسبة المجتمع P هي:

$$P\left[\hat{p}-Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}},\quad \hat{p}+Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right]$$

حيث $\sigma_{\hat{p}}=\sqrt{\frac{\hat{p}\hat{q}}{n}}$، و $\hat{q}=1-\hat{p}$

أما تقدير حجم العينة المطلوب لتقدير النسبة يكون على النحو التالي:

$n=\frac{1}{4}(\frac{Z_{\alpha/2}}{E})^2$ حيث e يمثل خطأ المعاينة.

<u>**مثال:**</u> كلية تحتوي على 500 =N طالب، أخذت عينة تتكون من

n=200 طالب، لتقدير نسبة الطلبة الحاصلين على درجة جيد، حيث

وجد أن ١٢ لديهم هذه الدرجة، المطلوب:

أ) أوجد ٩٠% مجال ثقة لنسبة الذين لديهم درجة جيد في هذه الكلية.

ب) عين الخطأ المطلق الأعظمي المرتكب عندما نفترض أن $\hat{p} = p$ بثقة ٩٧%.

ج) ما هو حجم العينة التي ينبغي دراستها لكي لا يتجاوز للخطأ في التقدير المقدار ٠,٠٣ و بثقة ٩٥%.

حل المثال: $x = 12, \quad n = 200, \quad \hat{q} = 0.94, \quad \hat{p} = \frac{12}{200} = 0.06$

من جدول التوزيع الطبيعي $Z_{\alpha/2} = 1,645 \quad , \quad 1-\alpha = 0,90, \alpha/2 = 0,05$

ويكون مجال ثقة ٩٠% لـ أ هو:

$$\left[0.06 - 1,645\sqrt{\frac{(0.06)(0.94)}{200}}, \quad 0.06 + 1,645\sqrt{\frac{(0.06)(0.94)}{200}}\right]$$

أي أن المجال يكون كالتالي [0.032; 0.087].

ب) لدينا $1-\alpha = 0.97$ ومن جدول التوزيع الطبيعي نجد $Z_{\alpha/2} = 2.17$

والخطأ المطلق الأعظمي المرتكب في تقدير أ هو :

$$e = \left| \pm 2.17\sqrt{\frac{(0.06)(0.94)}{200}} \right| = 0.036$$

ج) لدينا $Z_{\alpha/2} = 1,96$, $1-\alpha = 0,96, \alpha/2 = 0,025$ (من جدول التوزيع الطبيعي المعياري) وحجم العينة n التي ينبغي اختيارها هي:

$n = \frac{1}{4}(\frac{Z_{\alpha/2}}{E})^2 = \frac{1}{4}(\frac{1.96}{0.03})^2 = 1067.11$ إذن حجم العينة اللازم هو $n = 1067$

٧-٥- مجال الثقة للفرق بين نسبتين لمجتمعين $(P_1 - P_2)$:

عندما نريد المقارنة لنسب المجتمعات بالنسبة لصفة معينة، كأن نقارن نسبة النجاح في مدرستين، أو النسبة الطلاق في بلدين أو....... إذا كانت $x_1, x_2 ....x_n$ عينة عشوائية مسحوبة من مجتمع له توزيع برنولي $Bi\ (n_1, P_1)$ ، وعينة أخرى $y_1, y_2 ....y_n$ مسحوبة من مجتمع كذلك له توزيع برنولي، والعينيتين مستقلتين، وإذا أخذنا n_1 حجم العينة الأولى ، وn_2 حجم العينة الثانية، وكانت x_1 و x_2 عدد مرات النجاح في العينيتين على الترتيب، فإن مقدري P_1 وP_2 على التوالي:

$\hat{p}_1 = \frac{x_1}{n_1}; \hat{p}_2 = \frac{x_2}{n_2}$ ويكون $\hat{p}_1 - \hat{p}_2$ مقدرا غير متحيز للفرق $P_1 - P_2$ لأن $E(\hat{p}_1 - \hat{p}_2) = E(\hat{p}_1) - E(\hat{p}_2) = p_1 - p_2$

وباعتبار أن حجم العينيتين n_1 , n_2 كبيرتين، ولهما التوزيع الطبيعي وحسب نظرية النهاية المركزية، فإن المتغير العشوائي $(\hat{p}_1 - \hat{p}_2)$ له

تقريبا التوزيع الطبيعي بمتوسط $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$ و تباين:

$\sigma^2_{\hat{p}_1-\hat{p}_2} = v(\hat{p}_1 - \hat{p}_2) = V(\hat{p}_1) + V(\hat{p}_2) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$ ويكون المتغير

العشوائي كالتالي: $Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \rightarrow N(0,1)$

وباستخدام نفس الخطوات في الفقرات السابقة نجد أن مجال الثقة

$100(1-\alpha)\%$ للفرق بين النسبتين $P_1 - P_2$ هو كالتالي:

$$P\left[(\hat{p}_1 - \hat{p}_2) - Z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < P_1 - P_2 < (\hat{p}_1 - \hat{p}_2) + Z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}\right] = 1 - \alpha$$

مثال :

في استطلاع لرأي أجرته إحدى الجرائد المحلية حول فوز مرشح في الانتخابات البلدية، فسحبت عينة عشوائية حجمها ٥٠٠٠ شخص من الضاحية الشمالية ووجد أن ٤٥٠٠ يؤيدون هذا المشرح، وسحبت عينة أخرى من الناحية الجنوبية حجمها ٢٠٠٠ شخص ووجد أن ١٦٠٠ يؤيد هذا المرشح ، أوجد مجال ثقة ٩٠% للفرق بين نسبتي المؤيدين في الضاحيتين.

حل المثال:

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{4500}{5000} = 0.9, \quad \hat{q}_1 = 0.1, \hat{p}_2 = \frac{x_2}{n_2} = \frac{1600}{2000} = 0.8, \quad \hat{q}_2 = 0.2$$

من جدول التوزيع الطبيعي $Z_{\alpha/2} = 1,645 \quad , \quad 1-\alpha = 0,90, \alpha/2 = 0,05$

المعياري

وباستخدام النظرية السابقة يمكن ايجاد $P_1 - P_2$:

$$P\left[(0.1) - 1,645\sqrt{\frac{(0.9)(0.1)}{5000} + \frac{(0.8)(0.2)}{2000}} < P_1 - P_2 < (0.1) + 1,645\sqrt{\frac{(0.9)(0.1)}{5000} + \frac{(0.8)(0.2)}{2000}}\right] = 0,90$$

$$P(0.083 < p_1 - p_2 < 0.116) = 0.9$$

ومنه نستنتج أن نسبة المؤيدين لهذا المرشح في الضاحية الشمالية أكبر من نسبة المؤيدين في الضاحية الجنوبية.

تمارين: توزيعات المعاينة– التقديرات:

ت١ : تخضع أوزان الأطفال عند الولادة في مدينة ما للتوزيع الطبيعي بوسط ٣,٤ كغ وبتباين ١,٤٤ كغ، سحبت عينة عشوائية حجمها ٥٠ طفلا. المطلوب: ١) إيجاد توزيع المعاينة لهذه العينة.

٢) إيجاد احتمال أن يزيد متوسط الأوزان عن ٣,٧ كغ.

٣) إيجاد احتمال أن يكون الوزن ما بين (٣,٩ و ٣كغ).

ت٢: سحبت عينة عشوائية حجمها ٦٠٠ تلميذ في إحدى المدارس الابتدائية، فإذا كانت نسبة النجاح هي ٨٠%، أوجد:

- توزيع المعاينة لهذه العينة؟
- احتمال أن يزيد معدل النجاح عن ٨٢%؟
- احتمال أن يكون معدل النجاح ما بين ٧٧% و ٨٤% ؟

ت٣: مجتمع يتوزع طبيعيا، و σ^2 معلومة ،أوجد مستوى الثقة للمجالات التالية:

1) $(\bar{x} - (1.5)\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + (1.5)\frac{\sigma}{\sqrt{n}})$

2) $(\bar{x} - (2.5)\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + (2.5)\frac{\sigma}{\sqrt{n}})$

3) $(\bar{x} - (2.75)\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + (2.75)\frac{\sigma}{\sqrt{n}})$

ت٤: تم اختيار مجموعتين من الطلبة في مقياس الرياضيات، وسحبت عينيتين من طلاب وطالبات لإحدى المدارس وكانت النتائج كما يلي:

إذا علمنا أن معدات التحصيل تتبع التوزيع الطبيعي وأن لها نفس التباين

	طلاب	طالبات
الوسط الحسابي	١٣,٥	١٠,٥
الانحراف المعياري	٢,٥	٢
حجم العينة	١٠	١٢

أوجد ٩٩% مجال ثقة للفرق بين

$\mu_1 - \mu_2$ حيث μ_1 معدل المقياس

لدى الطلاب و μ_2 معدل المقياس لدى الطالبات، هل هناك فرق حقيقي بين المتوسطتين؟.

٧- اختبار الفروض الإحصائية test des hypothèses statistiques:

تمهيد: عند قيام الباحث بإجراء بحث ما، وبعد اختياره لعينة واستخلاصه لنتائج، فإنه سيكون في حالة شك من هذه النتيجة، هل هي راجعة إلى مجرد الصدفة أم ظاهرة حقيقية في المجتمع الأصلي. فاختبار الفرضيات يوفر للباحث تكرار التجربة عدة مرات من خلال عدة عينات، والتأكد من

النتائج التي حصل عليها، وكذلك يوفر له اجهد والنفقات، فالسر في اختبار الفرضيات يكمن في أن نقرر ما إذا كانت القيمة المعلنة (المختبرة) المعلمة المجتمع مثل متوسط المجتمع يجب أن يوافق عليها بالفعل مثل ما أنها مقبولة في ظاهرها، وذلك استنادا لما ستقدمه العينة من شواهد.

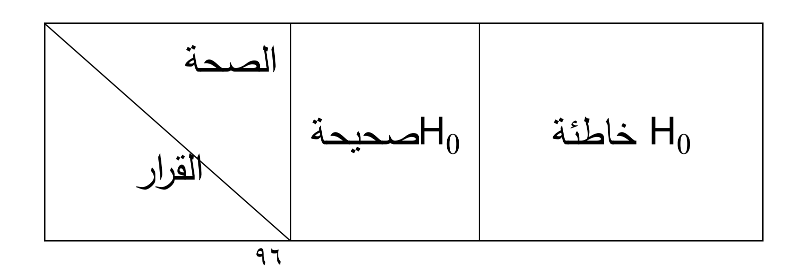
تعريف: الفرضة الإحصائية عبارة عن ادعاء أو تصريح (قد يكون صائبا أو خاطئا) حول معلمة أو أكثر لمجتمع أو لمجموعة من المجتمعات وعادة ما يصاغ الفرض الإحصائي يشكل عدم وجود اختلاف أو عدم وجود علاقة، وتسمى بفرضية العدم أو الصفرية hypothèse nulle ونرمز لها بـ (H_0)، هذه الفرضية هي التي تنطلق منها ولا نرفضها إلا إذا توفرت بيانات من العينة نقود إلى رفضها وهي التي تكون موضع الاختبار.

مثال: رمي قطعة نقود، قبل أن نلقي هذه القطعة نريد أن تكون متوازنة يعني أننا نريد اختبار الفرض أن المعلمة أ تساوي ٠,٥ فلو ألقينا ٣٠٠ مرة وحصلنا على الصورة ١٥٠ مرة فإننا نقبل بهذه الفرضية، وإذا حصلنا على ١٤٨ أو ١٥٢ مرة في ٣٠٠ رمية، فإن هذه النتيجة لا تقودنا إلى رفض الفرضية، أما إذا كان عدد الصور ضئيلا جدا نتحصل على ٢٠

صورة و ٢٥٠ مثلا فإننا نشك أن هذه القطعة غير متوازية عندئذ نرفض الفرضية، لأن احتمال وقوع القطعة متوازية ضئيل جدا.

ماذا لو أننا تحصلنا على ١١٥ صورة أو ١٧٢ صورة؟ إن على الباحث أن يجد طريقة جيدة لاتخاذ قرار القبول ($P= 0.5$) أو الرفض إذن علينا أن ندرك أن رفض الفرضية معناه أننا قررنا أنها خاطئة، بينما قبولنا لهذه الفرضية أننا يعني أننا لم نجد الأسباب الكافية لرفضها، عندئذ على الباحث أي يضع الفرضية المضادة أو البديلة hypothèse alternative، وهذه الأخيرة تكون صحيحة عندما نرفض الفرضية الصفرية.

وتأخذ الفرضية البديلة ثلاثة أشكال ($\theta>\theta_0 , \theta<\theta_0$)(اختبار أحادي Unilatéral) أو $\theta_0\neq\theta$(اختبار ثنائي) (bilatéral) عند استعمال أي اختبار لا يستبعد أن نرتكب أحد الخطأين، الخطأ الأول هو رفض الفرضية H_0 مع أنها صحيحة، والخطأ الثاني وهو قبول H_0 مع أنها خاطئة، والجدول التالي يبين الخطأين:



الصحة / القرار	H_0 صحيحة	H_0 خاطئة

خطأ من النوع الثاني	قرار جيّد	قبول H_0
قرار جيّد	خطأ من النوع الأول	رفض H_0

لتوضيح هذا الجدول نأخذ مثال حالة مريض مع إلزامية تناوله إلى الدواء:

في صحّة جيّدة	في صحّة غير جيّدة	حالة المريض / الدخول إلى المستشفى
β	قرار صائب	تناول الدّواء
قرار صائب	α	عدم تناول الدّواء

- الخطأ عن النوع الأول (α) هو عدم تناول الشخص للدواء مع أنه مريض، وهي كارثة بالنسبة له لأنها قد تؤدي به إلى الموت.

- للخطأ من النوع الثاني (β) وهو تناول الشخص للدواء وهو في صحة جيدة، وهي لا تقل خطورة عن الأولى.

$1-\beta$ يدعى بقوة الاختبار (La puissance du test) وهو رفض H_0 مع العلم بأنها خاطئة.

<u>٨-١- اختبارات حول المتوسطات:</u>

نستفيد فيما تطرقنا إليه في الفقرات السابقة (توزيع المعاينة والتقدير في تعيين مجالات الثقة، وسنتطرق إلى إجراء الاختبارات التالية:

٨-١-١- اختبار حول متوسط مجتمع طبيعي تباينه σ^2 معلوم:

ليكن لدينا ($x=(x_1, x_2 \ldots\ldots\ldots x_n$ عينة عشوائية مسحوبة من مجتمع طبيعي $N(\mu,\sigma^2)$ لها متوسط $(\bar{x})$ ، $\bar{x} \to N(\mu,\sigma^2)$

نريد اختبار الفرضية $H_0 : \mu = \mu_0$ ضد الفرضية البديلة H_1، التي تأخذ الأشكال التالية:

(اختبار من طرفين) $H_1 : \mu \neq \mu_0$

(اختبار من طرف واحد من جهة اليمين) $H_1 : \mu > \mu_0$

(اختبار من طرف واحد من جهة اليسار) $H_1 : \mu < \mu_0$

ولدينا عينة كبيرة الحجم مع σ^2 معلوم، فإحصاءة الاختبار تكون باستعمال المتغير العشوائي (الإحصاءة) التالي: $Z_0 = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

والفرضية الصفرية تكون كما يلي: $H_0 : \mu = \mu_0$

والمناطق الحرجة لهذا الاختبار (توزيع طبيعي و σ معلومة، حجم كبير).

رفض H_0 إذا	الفرضية البديلة
$Z_0 < -Z_\alpha$	$\mu < \mu_0$
$Z_0 > Z_\alpha$	$\mu > \mu_0$

$\mu \neq \mu_0$	$Z_0 > Z_{\alpha/2}$ أو $Z_0 < -Z_{\alpha/2}$

ملاحظة: يمكن استخدام التوزيع الاحتمالي الطبيعي في اختبار قيمة مفترضة لمتوسط المجتمع طالما $n \geq 30$ (نظرية النهاية المركزية) وكذلك يمكننا استخدام هذا التوزيع في حالة 30 > n شرط أن المجتمع يتوزع طبيعيا وكانت σ^2 معلومة.

مثال: أظهرت نتائج إحصاء التعداد السكاني لإحدى الدول أن متوسط حجم الأسرة بلغ ٠٥ أفراد والتباين بلغ ٠٤ أفراد، أخذت عينة من ٥٠٠ أسرة من كل أنحاء البلاد، ووجد أن متوسط حجم الأسرة بلغ ٠٦ أفراد، المطلوب هل يمكننا القول أن متوسط حجم الأسرة بلغ ٠٦ أفراد، المطلوب هل يمكننا القول أن متوسط الأسرة هو أكبر من بيانات التعداد السكاني عند مستوى المعنوية (٠١%).

حل المثال: لنشكل الفرضيات التالية: $H_0 : \mu = 5, \quad H_1 : \mu > 5$

من جدول التوزيع الطبيعي: $Z_{\alpha/2} = 2.33;;; \quad \alpha = 0.01 \quad ,$

$$Z_0 = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{6-5}{2/\sqrt{500}} = 11.18$$

نلاحظ أن قيمة z المحسبة تقع في منطقة الرفض H_0 نرفض الفرضيةH_0 وبالتالي نقبل بالفرضية البديلة (H_1) أي أن متوسط حجم الأسرة في هذا البلد أكبر من ٠٥ أفراد).

٨-١-٢- اختبارات حول متوسط مجتمع طبيعي تباينه σ^2 مجهول:

ذكرنا في فقرات سابقة عندما يكون تباين المجتمع σ^2 غير معلوم ويكون توزيع العينة طبيعيا، فإن الإحصاءة الاختبار تكون باستخدام توزيع t بـ (n−1) درجة حرية، حيث $T_0 = \frac{\bar{x} - \mu}{s / \sqrt{n}}$ ، S^2 تباين العينة، ونستخدم نفس خطوات التي استخدمناها في الفقرة السابقة، ونبين في الجدول التالي المناطق الحرجة.

المناطق الحرجة لاختبار $H_0 : \mu = \mu_0$ (توزيع طبيعي، σ^2 مجهولة).

رفضH_0 إذا	الفرضية البديلة
$t_0 < -t_\alpha$	$\mu < \mu_0$

$\mu > \mu_0$	$t_0 > t_\alpha$
$\mu \neq \mu_0$	$t_0 > t_{\alpha/2}$ أو $t_0 < -t_{\alpha/2}$

<u>مثال:</u> في مدرسة ابتدائية كانت أوزان التلاميذ تخضع للتوزيع الطبيعي بمتوسط ٣٥ كغ، طرأ نظام جديد لهذه المدرسة بحيث تم إدخال الوجبة الصباحية، وبعد مدة أراد باحث اختبار نجاعة هذا البرنامج الغذائي، فأخذ عينة مكونة من ٢٠ طفلا، وتم وزنهم فكان متوسط الوزن ٣٨ كغ بانحراف معياري ٠٥ كغ، فهل هذا البرنامج أثبت نجاعته أم لا عند مستوى معنوية (٠٥%).

<u>حل المثال:</u>

$$H_0 : \mu = 35$$
$$H_1 : \mu > 35$$

من جدول توزيعt

$$, \quad \alpha = 0.05 \quad t_{\alpha,n-1} = t(0.05,19) = 1.729;$$

$$T_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{38-35}{5/\sqrt{20}} = 2.68$$

نلاحظ أن t المحسوبة تقع في منطقة الرفض، وبالتالي نستنتج أن البرنامج الغذائي المقدم لهؤلاء التلاميذ قد أثبت (نجاعته):

٨-١-٣- اختبارات حول الفرق بين متوسطي مجتمعين طبيعيين تباينهما معلومان:

عندما نريد المقارنة بين متوسطي مجتمعين، فإننا سنختبر أن الفرق يرجع إلى الصدفة (أي لا يوجد اختلاف بينهما)، أم أن هذا الفرق جوهري (يوجد اختلاف بينهما).

لدينا مجتمعين وسيطيهما الحسابيين μ_1 و μ_2 وتباينهما σ_1^2 و σ_2^2 على الترتيب، سحبنا عينيتين n_1 و n_2 من هذين المجتمعين فكان متوسطيهما الحسابيين $\bar{x}_1$ و $\bar{x}_2$ على الترتيب (حيث أنهما مستقلين عن بعضهما البعض) فإحصاءة الاختبار تكون كالتالي:

$$Z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

وتكون صياغة الفرضيات كما يلي: $H_0 : \mu_1 = \mu_2$ ضد الفرضية البديلة التي تأخذ نفس الأشكال السابقة اختبار من طرفين $H_1 : \mu_1 \neq \mu_2$

اختبار من طرف واحد من جهة اليمين $H_1 : \mu_1 > \mu_2$

اختبار من طرف واحد من الجهة اليسار $H_1 : \mu_1 < \mu_2$

٨-١- ٤- اختبارات حول الفرق بين متوسطي مجتمعين طبيعيين تباينهما مجهولان:

في حالة ما إذا كان حجم العينيتين ٣٠ > n_2 و n_1 وتباين المجتمعين σ_1^2 و σ_2^2 مجهولين، سنتطرق إلى نفس الحالة التي استخدمناها في تقدير مجال الثقة.

الحالة ١: تساوي تباين المجتمعين $\sigma_1^2 = \sigma_2^2 = \sigma^2$:

فندرس هذه الحالة وكأن للمجتمعين تباين مشترك $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ونستخدم إحصاءةT حيث :

$$t_0 = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{SP\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

حيث t هو توزيع ستيودنت بـ $V = n_1 + n_2 - 2$ درجة حرية و

$$SP^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

، حيث S_1^2 تباين العينة في المجتمع الأول S_2^2 تباين العينة في المجتمع الثاني، وبقية الخطوات هي نفسها كما استخدمناها في الفقرات السابقة.

مثال: سحبت عينة عشوائية من بعض الأقسام النهائية في ثانوية A (تتبع نظام الدروس الإضافية لدى تلاميذها) وسحبت عينة عشوائية مستقلة عن الأولى من إحدى الأقسام النهائية في ثانوية B (لا تتبع نظام الدروس الإضافية لدى تلاميذها) أراد باحث دراسته مستوى التحصيل في الشعب العلمية، فاختار عينة حجمها $n_1= 13$ من ثانوية A وعينة أخرى حجمها $n_2= 14$ من ثانوية B، (علما أن معدلات التحصيل في كلا الثانويتين تتبعان التوزيع الطبيعي).

وفي نهاية الفصل أعطى نفس الامتحان لكلا التلاميذ في الثانويتين، فكانت النتائج كما يلي: $S_2^2= 2$, $S_1^2= 4$,

$\bar{x}_1 = 12.75, \quad \bar{x}_2 = 12$ ، فهل أن برنامج الدروس الإضافية أدى إلى نتائج جيدة أم لا؟ عند مستوى معنوية ٥٠%.

حل المثال:

لنختبر صحة الفرضية التالية:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

(من جدول توزيع ستيودنت)

$$1-\alpha = 0,95, \Rightarrow \quad \alpha/2 = 0,025, \quad t_{\alpha/2} = t(\alpha/2 \quad , \quad n_1+n_2-2) = t(0.025, 25) = 2.060$$

$$t = \frac{(\overline{x}_1 - \overline{x}_2)}{SP\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{12.75-12}{1.72\sqrt{\frac{1}{13}+\frac{1}{14}}} = 1.13 \text{ ،} \quad SP^2 = \frac{(12)4+(13)2}{25} = 2.96$$

بملاحظة $-2.06 < t = 1.13 < 2.06$ (t المحسوبة تقع في منطقة قبول H_0 أي أنه لا يوجد اختلاف جوهري بين نتائج الثانويتين من خلال تقديم الدروس أو غير تقديمهما.

<u>الحالة ٢: عدم تساوي تباين المجتمعين</u> $\sigma_1^2 \neq \sigma_2^2$:

في حالات عديدة لا يعقل دائما أن نفترض تساوي تباين المجتمعين (عندما يكون تباين المجتمعين غير معلومين في هذه الحالة نستخدم إحصاءة الاختبار التالية:

$$T^* = \frac{(\overline{x}_1 - \overline{x}_2) - \Delta_0}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

، يتوزع تقريبا بتوزيع t مع درجات حرية مقدمة بالصيغة التالية:

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2 / n_1)^2}{n_1 - 1} + \frac{(S_2^2 / n_2)^2}{n_2 - 1}}$$

٨-٢- اختبار الفرضيات حول نسبة في مجتمع:

هناك حاجة لاختبار فرضية معينة حول نسبة حدث معين، بدلا من اختيار الفرضية حول متوسط حدث في مجتمع، ويحدث ذلك في عدة ظواهر خاصة التي لا يمكن حساب متوسطاتها، مثلا نسبة الطلاق، نسبة الزواج، الأمية، البطالة.......إن المقدر $\hat{p}$ هو مقدر غير متحيز بحيث $\hat{p} = \frac{x}{n}$ (x: عدد النجاحات ، n هو حجم العينة في الحالة التي يكون فيها حجم كبير $n \geq 30$).

إن خطوات الاختبار لنسبة واحدة هي : $H_0: P = P_0$

وتكون الفرضية البديلة بالأشكال التالية: $H_1: P_1 \neq P_0$

$H_1: P_1 > P_0$

$H_1: P_1 < P_0$

أما إحصاءة الاختبار فهي : $Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{x - np_0}{\sqrt{np_0 q_0}}$ ،

$q_0 = 1 - p_0$

٨-٣- اختبارات حول الفرق بين نسبتي مجتمعين:

إذا أردنا المقارنة بين نسبتي في مجتمعين نتبع نفس خطوات الاختبار بالنسبة الفرق بين وسطين، حيث تكون صياغة الفرضية الصفرية كالتالي: $H_0: P_1 = P_2$ (أن تساوي نسبتي المجتمعين، بحيث أنهما تتبعان توزيع برنولي) ضد الفرضية البديلة التي تأخذ الأشكال التالية:

$H_1: P_2 \neq P_1$

أو$H_1: P_1 > P_2$

أو$H_1: P_1 < P_1$

وقد تطرقنا في الفصل السابق عندما يكون الحجم الكبير للعينيتين $n_1, n_2 \geq 30$ ، فإن التغير العشوائي Z يكون كالتالي

$Z_0 = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$ ، أي تقريبا له توزيع طبيعي معياري N

(0,1)، حيث $\hat{p}_1 = \frac{x_1}{n_1}$ و $\hat{p}_2 = \frac{x_1}{n_1}$

إذا أخذنا بعين الاعتبار صحة الفرضية H_0 أي $P_1 = P_2 = P$ فتصبح

علاقة Z كالتالي: $Z_0 = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{pq(\frac{1}{n_1} + \frac{1}{n_2})}}$ ، وبما أن $\hat{p}$ هو مقدر غير متحيز

حيث $\hat{p} = \frac{\bar{x}_1 + \bar{x}_2}{n_1 + n_2}$ x_1 و x_2 عدد النجاحات في العينيتين الأولى والثانية

على الترتيب وبالتالي نحصل على العلاقة النهائية ل Z. وهي كما يلي:

$Z_0 = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}}$ ، $\hat{q} = 1 - \hat{p}$

<u>**مثال:**</u> قامت إحدى الجمعيات المدنية المسمات "**سلامتك**" بدراسة حول

تطبيق حزام الأمان عند السائقين في الطرقات فاختارت عينيتين

عشوائيتين مستقلتين في ولايتين A و B قدرها ١٠٠ سائق من كل ولاية ،

فتبين أن ٨١ و ٧٣ سائق يطبقون استعمال الحزام في الولايتين على

الترتيب، فهل تقدم هذه المعلومات الإحصائية دلالة كافية أن B عند مستوى معنوية ٠٥%.

حل المثال: نختبر صحة الفرضية :

$$H_0 : P_1 = P_2 ; H_1 : P_1 \neq P_2$$

$$n_1 = 100, \quad x_1 = 81, \quad \hat{p}_1 = \frac{x_1}{n_1} = \frac{81}{100} = 0.81$$

$$n_2 = 100, \quad x_1 = 73, \quad \hat{p}_2 = \frac{x_1}{n_1} = \frac{73}{100} = 0.73$$

$$\hat{p} = \frac{\bar{x}_1 + \bar{x}_2}{n_1 + n_2} = \frac{81 + 73}{200} = 0.77, \quad \hat{q} = 0.23$$

$$Z_0 = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.81 - 0.73}{\sqrt{(0.77)(0.23)(\frac{2}{100})}}$$

$$1 - \alpha = 0,95, \Rightarrow \quad Z_{\alpha/2} = 1.96$$

نلاحظ أن Z المحسوبة تقع ضمن منطقة قبول H_0، نستنتج أن نسبة التطبيق متساوية أي لا يوجد فرق جوهري بين نسبة التطبيق الحزام في الولايتين.

تمارين: اختبار الفرضيات:

ت١: في كل حالة من الحالات التالية: حدد صحة صياغة الفرضيات:

$a): H_0 : \mu = 30; H_1 : \mu \neq 30$
$b): H_0 : \overline{x} = 70; H_1 : \overline{x} \neq 70$
$c): H_0 : P = 0.3; H_1 : P_1 = P_2$

ت٢:

نريد دراسة شدة المقاومة في أسلاك معدنية مصنوعة من خليط من معدنين أحدهما نفيس، علما أن توزيع شدة المقاومة تتبع التوزيع الطبيعي بانحراف معياري $\sigma = 50\Omega$ ، فأخذنا عينة حجمها ١٦، فوجد أن متوسط شدة المقاومة يقدر بـ $\overline{x} = 2250\Omega$ اختبر الفرضية عند المتوسط العام لشدة المقاومة والذي يقدر بـ 2500Ω ، عند مستوى معنوية ٠٥% ؟

ت٣:

شركتين A و B تعملان في إطار الصناعات المطاطية المختصة في إنتاج عمليات السارات، أراد باحث دراسة مدى قدرة تحمل العجلات للاحتكاك، فاختار عينتين عشوائيتين مستقلتين حجمها ٢٠ عجلة من كل شركة، وثم اختبار العجلات مقاسة (mg/1000 cycles) ، وتم قياس

متوسط والانحراف المعياري لتحمل العجلات، فكانت النتائج كما يلي:

$\bar{x}_1 = 18$, $\bar{x}_2 = 14$, $S_2 = 6$.

المطلوب: هل توجد دلالة كافية لاختلاف متوسط التحمل بالنسبة لقدرة العجلات عند مستوى معنوية $\alpha = 0.05$ باعتبار أن متوسط التحمل للمجتمعين يتبعان التوزيع الطبيعي مع اختلاف تبايناهما.

٩- اختبار كي مربع test de Khi deux:

سنتطرق فقط إلى اختبار كي تربيح للاستقلالية علما أنه توجد عدة اختبارات بالنسبة لكي مربع وهي اختبار المقارنة بين توزيعات متعددة الحدود، اختبار جودة التوفيق (تحديد المعلمة/ عدم تحديد المعلمة).

وتكمن أهمية اختبار كي تربيع للاستقلالية في الكشف عن العلاقة بين متغيرين x و y (جداول التوافق)، ليكن لدينا متغيرين عشوائيين x و y يأخذ القيم التالية $\{x_1......x_i........x_2\}$ و $\{y_1......y_i........y_c\}$ مع الاحتمالات $P(x= x_i)= P_i$ و $P(y= y_j)= q_j$، نقول أيضا مستقلة إذا وفقط إذا كان:

$$P\{ x= x_i , y=y_j\} = P (x= x_i) P (y= y_i) = P_i q_j$$

$$i=1....n, \quad j=1....m \quad , \quad P_i = p(x=i) = \sum_{j=1}^{m} P_{ij} \quad , \quad i=1....n$$

$$q_j = p(y=j) = \sum_{i=1}^{n} P_{ij} \quad , \quad j=1....m$$

صياغة الفرضيات:

$$H_0 : P_{ij} = p_i q_j \quad \text{لكل} \quad i=1....n, \quad j=1....m$$
$$H_1 : P_{ij} \neq p_i q_j \quad \text{لبعض} \quad i=1....n, \quad j=1....m$$

لاختبار الاستقلالية بين المتغيرين نستخدم اختبار كي تربيع ولتوضيح هذا الاختبار نأخذ المثال التالي:

مثال: لاختبار فعالية لقاح معين (x) ضد الإنفلونزا العادية، ثم اختيار ٥٠٠ شخص مصاب بالأنفلونزا العادية وثم لقاحهم بلقاح (x) وقد قسم العمر إلى ثلاث فئات (أقل من ٣٠ سنة) (٣٠-٥٠ سنة)، (أكثر من ٥٠ سنة) ، وبعد فترة ثم فحصهم من جديد كانت البيانات مبينة في الجدول التالي:

الحالة / فئات العمر	شفى من الإصابة	لم يشفى من الإصابة	المجموع
أقل من ٣٠ سنة	١٢٠	٨٠	٢٠٠
٣٠-٥٠ سنة	٥٠	١٠٠	١٥٠
أكثر من ٥٠ سنة	٢٠	١٣٠	١٥٠
المجموع	١٩٠	٣١٠	٥٠٠

هل يمكن الاستنتاج أن فعالية اللقاح تختلف بين فئات العمر عند مستوى $\alpha = 0.05$ ؟

حل المثال:

١) صياغة الفرضيات

H_0: فعالية اللقاح مستقل عن العمر (لا توجد علاقة)

H_1: فعالية اللقاح غير مستقل عن العمر (توجد علاقة)

١) حساب التكرار المتوقع $E_{ij} = \frac{R_i C_j}{n}$

أقل من ٣٠ سنة	٧٦	١٢٤
٣٠-٥٠ سنة	٥٧	٩٣
أكثر من ٥٠ سنة	٥٧	٩٣

حساب قيمة كي مربع وفقا للعلاقة : $\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$

$$\chi^2 = \frac{(120-76)^2}{76} + \frac{(80-124)^2}{124} + \frac{(50-57)^2}{57} + \frac{(100-93)^2}{93} + \frac{(20-57)^2}{57} + \frac{(130-93)^2}{93}$$

$$= 25,47 + 15,61 + 0,86 + 0,67 + 24,01 + 14,72 = 81,34$$

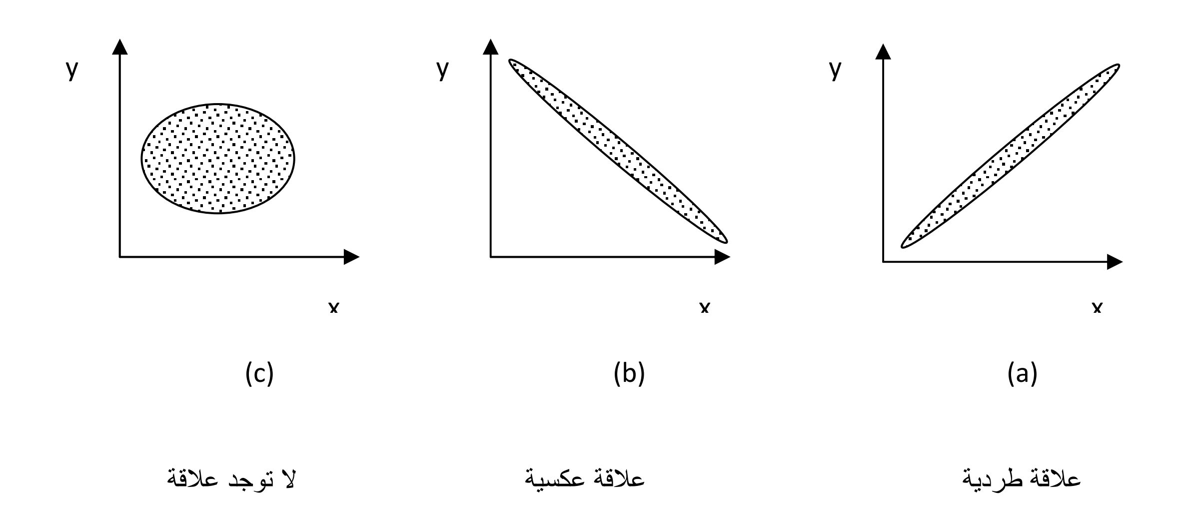
$$v = (r-1)(c-1) = (2)(1) = 2, \quad \chi_{0.05,2} = 5,99$$

بما أن قيمة كي تربيع تقع في منطقة رفض H_0، لذا يتم رفض H_0 أي أن فعالية اللقاح غير مستقل عن فئات العمر (توجد علاقة عند مستوى $\alpha = 0.05$ ، فالأشخاص كبار السن فعالية اللقاح هي أقل من فئات الشباب وهكذا.

١٠- الارتباط والانحدار البسيط بين متغيرين:

تمهيد: عند تحليلنا لبعض الظواهر اجتماعية كانت أم اقتصادية....لا نكتفي بالقول بأن هناك علاقة أم عدم وجودها (تحليل نوعي)، بل نرغب كذلك بإجراء دراسته كمية لها، مثلا لدينا متغيرين (x و y) متى نقول أنهما يتغيران في اتجاه واحد؟ (علاقة طردية) ، ومتى نقول أنهما يتغيران عكسيا؟ كيف يمكننا قياس قوة الارتباط.

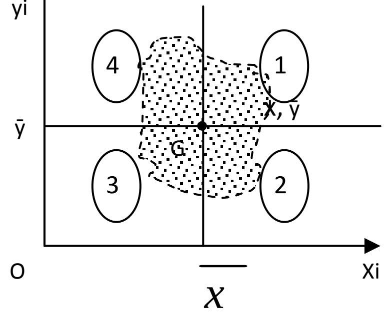
إذا كان هناك ارتباط بين x و y ، سنمثل كل مشاهدة i بنقطة إحداثيات (x_i, y_i) في معلم كارتيزي، لنلاحظ سحابة النقاط التالية:



(a) علاقة طردية — (b) علاقة عكسية — (c) لا توجد علاقة

١٠-١- التباين المشترك (التغاير) La covariance:

نحاول في هذه الفقرة البحث عن كمية تحدد وجود ارتباط بين المتغيرين ونقيس قوته ونعطي اتجاهه، لنأخذ الشكل البياني التالي: الذي يبين سحابة نقاط، لنأخذ مركز ثقل السحابة (le centre de gravité du nuage de points)



(المتوسط G) ، بإجراء عملية انسحاب المعلم translation بوضع مركز على النقطة (G) نحصل

على أربع مناطق.

لنعتبر الكمية التالية : $\alpha_i = (x_i - \overline{x})(y_i - \overline{y})$ نتحصل على أربع حالات :

١) بالنسبة للمنطقة (١) الجداء α_i يكون موجبا لأن النقطتان $(x_i - \overline{x})$ و $(y_i - \overline{y})$ موجبتين.

٢) المنطقة (٢) الجداء α_i سالب، ٣) المنطقة (٣) الجداء α_i موجب.

٤) المنطقة (٤) الجداء α_i سالب.

لنهتم الآن جمالي السحابة ونهتم بالكمية $\beta = \frac{1}{n}\sum_{i=1}^{n}\alpha_i$ لنرجع إلى تفسير الأشكال السابقة (c,b,a).

بالنسبة للشكل (a) أغلب النقاط موجودة في المنطقة (١) و (٣)، إذن علاقة طردية للشكل (b) أغلب النقاط موجودة في المنطقة (٢) و (٤)، إذن هناك علاقة عكسية.

بالنسبة للشكل (c) تشتت النقاط موزعة على النقاط الأربع تقريبا مما يجعلنا نتوقع قيمة β قريبة من الصفر (لا توجد علاقة). إن قيمة β تعبر عن الارتباط بين المتغيرين x و y، وتعبر عن شدته وعن اتجاهه فخلاصة القول أن قيمة β تدعى بالتباين المشترك بين المتغيرين x و y ونرمز له بـ cov (x,y)، والذي يحسب بالصيغة التالية:

$\operatorname{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ وهذه الحالة تكون فيها كل التكرارات مساوية للواحد أو الصفر فقط (كل نقطة على البيانات لا تمثل إلا فردا واحدا من المجتمع (الحالة الخاصة).

<u>مثال١:</u> <u>لدينا</u> الجدول التالي الذي يمثل الوزن (x) والطول (y):

(kg) x_i	٦٠	٨٠	٥٥	٦٥	٤٠
(cm) y_i	١٦٠	٢٠٠	١٥٠	١٤٠	١٣٠

<u>المطلوب:</u> حساب cov (x,y) بين الطول والوزن.

<u>حل المثال١:</u>نستخدم الجدول التالي بغية تسهيل الحساب:

	x_i	y_i	$x_i\ y_i$	x_i^2	y_i^2
	60	١٦٠	٩٦٠٠	٣٦٠٠	٢٥٦٠٠
	٨٠	٢٠٠	١٦٠٠٠	٦٤٠٠	٤٠٠٠٠٠
	٥٥	١٥٠	٨٢٥٠	٣٠٢٥	٢٢٥٠٠
	٦٥	١٤٠	٩١٠٠	٤٢٢٥	١٩٦٠٠
	٤٠	١٣٠	٥٢٠٠	١٦٠٠	١٦٩٠٠
المجموع	٣٠٠	٧٨٠	٤٨١٥٠	١٨٨٥٠	١٢٤٦٠٠

$$\bar{x} = \frac{\sum x_i}{n} = \frac{300}{5} = 60, \quad \bar{y} = \frac{\sum y_i}{n} = \frac{780}{5} = 156$$

$$\operatorname{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} = 9630 - (60)(156) = 270$$

أما إذا لم يمكن أن نجمع عدة نقاط في موقع واحد، أي عندما تكون التكرارات n_{ij} أعداد كيفية تختلف عن (٠ أو 1) (الحالة العامة) فصيغة التباين المشترك تكون كما يلي:

$$\operatorname{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} x_i y_j - \bar{x}\bar{y}$$

<u>مثال٢:</u> أجريت دراسة في شركة ما حول الخبرة (x) والأجر (y) فأختبر ٥٠ عامل عشوائيا، وتم الحصول على البيانات التالية:

الأجر بوحدة نقدية / الخبرة بالسنوات	أقل من ٢٠٠ ون	٢٠٠– ٤٠٠ ون	أكثر من ٤٠٠ ون	المجموع
أقل من ٠٥ سنوات	٢	٢	٦	١٠
٠٥–١٠ سنوات	٢	٣	٨	١٣
أكثر من ١٠ سنوات	٥	٥	١٧	٢٧
المجموع	٩	١٠	٣١	٥٠

<u>المطلوب:</u>

حساب $\operatorname{cov}(x, y)$؟

حل المثال ٢ : نحسب مركز الفئات:

الأجر (y) / الخبرة (x)	١٠٠		٣٠٠		٥٠٠		المجموع
٢,٥	٢	٥٠٠	٢	١٥٠٠	٦	٧٥٠٠	٩٥٠٠
٧,٥	٢	١٥٠٠	٣	٦٧٥٠	٨	٣٠٠٠٠	٣٨٢٥٠
١٢,٥	٥	٦٢٥٠	٥	١٨٧٥٠	١٧	١٠٦٢٥٠	١٣١٢٥٠
المجموع		٨٢٥٠		٢٧٠٠٠		١٤٣٧٥٠	١٧٩٠٠

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_{i.} x_i = \frac{1}{50}((10*2.5)+(13*7.5)+(27*12.5)) = 9.2$$

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{p} n_{.j} y_j = \frac{1}{50}((9*100)+(10*300)+(31*500)) = 388$$

$$\text{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{p} x_i y_j - \bar{x}\bar{y} = \frac{1}{50}(179000) - (9.2)(388) = 10.4$$

١٠-٢- معامل الارتباط الخطي Le coefficient de corrélation linéaire

في الفقرة السابقة اعتبرنا أن التباين المشترك يعبر عن الارتباط بين متغيرين، وقد بينا أن التباين المشترك يبين اتجاه هذ الارتباط من خلال المناطق الأربعة، لكن السؤال المطروح متى نقول أن قيمة التباين المشترك كبيرة؟ ابتداء من أي قيمة يكون فيها قويا؟ ليكن لدينا مثلا زوجان من المتغيرات (x,y) و (s,t)، لنفترض أن قيمة =cov (x,y) 300وقيمة 0.9 =cov (s,t)، هل يمكن القول أن قيمة cov (x,y) أقوى من cov (s,t) باعتبار أن القيمة (300 < 0.9)؟

إذا أضفنا معلومة جديدة وهي أن x و y قد قيسا باللتر وأن s و t قيسا بالميليلتر، هل يبقى استنتاجنا صحيحا؟ إن قيمة التباين المشترك لها علاقة بوحدة قياس x وy ، إذا تغيرت وحدة وقياس x مثلا، فقيمة cov(x,y) حتما ستتغير.

إذن تحتاج إلى معامل آخر يقيس لنا قوة العلاقة وهو ما يعرف بالارتباط الخطى الذي يأخذ الصيغة التالية: $\rho = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}$،وهو محصور دائما بين (١) و (-١) ، $-1 \leq \rho \leq 1$ ، فكلما اقترب من هاتين

القيمتين تكون العلاقة قوية. اختلاف الإشارة يحدد العلاقة طردية كانت أم عكسية؟ أما إذا اقترب من الصفر فتكون العلاقة ضعيفة، وإذا كان $\rho = 0$ فلا توجد علاقة.

مثال: أحسب معامل الارتباط بالنسبة للمثالين السابقين.

حل المثال: بالنسبة لمعامل الارتباط الخاص بالمثال (١) يكون على النحو التالي:

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{n} - (\bar{x})^2} = 13.03, \quad \sigma_y = \sqrt{\frac{\sum y_i^2}{n} - (\bar{y})^2} = 24.16$$

إذن

$$\rho = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y} = \frac{270}{(13.03)(24.16)} = 0.85$$

بالنسبة لمعامل الارتباط الخاص بالمثال (٢)فيكون على النحو التالي:

نحسب σ_x و σ_y من خلال الصيغتين التاليتين:

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{k} n_{i.} x_i^2 - (\bar{x})^2 = \frac{1}{50}((10*6.25)+(13*56.25)+(27*156.25)) = 15.61, \quad \sigma_x = \sqrt{15.61} = 3.95$$

$$\sigma_y^2 = \frac{1}{n}\sum_{j=1}^{p} n_{.j} y_j^2 - (\bar{y})^2 = \frac{1}{50}((9*10000)+(10*90000)+(31*250000)) = 24256, \quad \sigma_y = \sqrt{24256} = 155.74$$

$$\rho = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y} = \frac{10.4}{(3.95)(155.74)} = 0.0169$$

١٠-٢-١- معامل الارتباط لبيرسون:

قلنا سابقا أن معامل الارتباط يقيس قوة الارتباط الخطية بين متغيرين كميين، وهو مؤشر إحصائي جد هام لقياس العلاقة بين متغيرين.

مثال: دراسة العلاقة بين الدخل والاستهلاك، والطول والوزن.......ويمكن إيجاده بالصيغة التالية (طريقة الانحرافات):

$$r_p = \frac{S_{xy}}{\sqrt{(S_{xx})(S_{yy})}}$$

حيث:

$$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$S_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y}$$

وتوجد عدة صيغ لحساب r_p نذكر منها:

$$r_p = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n(\bar{x})^2)(\sum y_i^2 - n(\bar{y})^2)}}$$

١٠-٢-٢- اختبار معنوية معامل الارتباط البسيط لبيرسون:

بعد تحديد قيمة r_p الذي يمثل معامل الارتباط لبيرسون بين أزواج المتغيرين x وy وممكن استخدامه أيضا في تقدير معامل الارتباط

للمجتمع (ρ)، ومن المفيد ايضا اختبار معنوية هذا المعامل وذلك في الحالات التالية:

تحديد مستوى المعنوية

١)
$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0 \, ou \quad \rho > 0 \quad ou \quad \rho < 0$$

فإحصاءة الاختبار $T_0 = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$، رفض H_0 إذا كان

$$|t| \geq t_{\alpha/2, n-2}$$

٢)
$$H_0 : \rho = \rho_0$$
$$H_1 : \rho \neq \rho_0 \, ou \quad \rho > \rho_0 \quad ou \quad \rho < \rho_0$$

مع حجم عينة ($n \geq 25$) نستخدم تحويلة فيشر

$$Z = \arctan h \quad r = \frac{1}{2} Ln \frac{1+r}{1-r}$$

حيث أنها تتوزع طبيعيا بوسط $\mu_z = \arctan h \quad \rho = \dfrac{1}{2} Ln \dfrac{1+\rho}{1-\rho}$ و بتباين

$$\sigma_z^2 = \frac{1}{n-3}$$

أما إحصاءة الاختبار فهي $Z_0 = \dfrac{Z_r - Z_{\rho_0}}{1/\sqrt{n-3}}$ علما أن

$$Z_r = \frac{1}{2} Ln \frac{1+r}{1-r}, Z_{\rho_0} = \frac{1}{2} Ln \frac{1+\rho_0}{1-\rho_0}$$

رفض H_0 إذا كان $|z| \geq z_{\alpha/2}$

٣) $H_0 : \rho_1 = \rho_2$
$H_1 : \rho_1 \neq \rho_2 \; ou \quad \rho_1 > \rho_2 \quad ou \quad \rho_1 < \rho_2$

أما إحصاءة الاختبار فهي $Z_0 = \dfrac{Z_{r1} - Z_{r2}}{\sqrt{\dfrac{1}{n_1 - 3} + \dfrac{1}{n_2 - 3}}}$، رفض H_0 إذا كان $|z| \geq z_{\alpha/2}$

مثال: أخذت عينة حجمها خمسة أسر وذلك بغية قياس العلاقة بين حجم الدخل وحجم الاستهلاك على السلع الضرورية ، فإذا كان معامل الارتباط $r = 0.6$.

المطلوب: هل ان قيمة معامل الارتباط البسيط المحسوب تدل على وجود علاقة معنوية بين المتغيرين وفقا لمعطيات العينة عند مستوى $\alpha = 0.01$؟

حل المثال:

١) صياغة الفرضية $H_0 : \rho = 0$
$H_1 : \rho \neq 0$

٢) $T_0 = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \dfrac{(0.6)\sqrt{3}}{\sqrt{1-(0.6)^2}} = 1.3, \quad t_{\alpha/2, n-2} = t_{0.005,3} = 5.841; \quad -t_{0.005,3} = -5.841$

الاستنتاج: بمأن قيمة T_0 تقع في منطقة قبول H_0 ، يعني لا توجد علاقة معنوية بين المتغيرين أي أن ($\rho \neq 0$) وفقا لبيانات العينة وعند مستوى معنوية $\alpha = 0.01$.

١٠-٢-٣- معامل ارتباط الرتب لسبيرمان (spearman) :

إذا كانت الظاهرة محل الدراسة تحتوي على متغيرين وصفيين ترتيبين، ومثال على ذلك قياس العلاقة بين تقديرات الطلبة في مادتين ، أو العلاقة بين درجة تفضيل المستهلك لسلعة معينة ، ومستوى الدخل، فإنه يمكن استخدام طريقة" بيرسون "السابقة في حساب معامل ارتباط يعتمد على رتب مستويات المتغيرين كبديل للقيم الأصلية ، ويطلق على هذا المعامل " معامل ارتباط سبيرمان" ويعبر عنه بالعلاقة التالية :

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

١٠-٢-٤- بعض الخصائص لمعامل ارتباط الرتب لسبيرمان:

يتصف هذا المعامل بالخصائص التالية:

- قيمة هذا المعامل تقع ضمن المجال $(-1 \le r_s \le 1)$، بحيث انه إذا كانت $6\sum d_i^2 = 0$ فان $r_s = 1$، و إذا كانت $\left(6\sum d_i^2 = 2n(n^2-1)\right)$ فان $r_s = -1$، و إذا كانت $\left(6\sum d_i^2 = n(n^2-1)\right)$ فان $r_s = 0$.
- قيمة هذا المعامل عند حساب العلاقة بين متغيرين كميين لا تساوي بالضبط قيمة معامل بيرسون، وذلك بسبب التعامل مع رتب القيم بدل القيم الاصلية .

١٠-٢-٥- الاستدلال حول معامل ارتباط الرتب لسبيرمان:

أ) اختبار الفرضيات:

أ-١) اختبار معامل ارتباط لسبيرمان (اختبار هوتلينغ بابستtese de hotelling- pabst)

إن اختبار الرتب لسبيرمان هو من الاختبارات اللامعلمتية ، ويستخدم في اختبار العلاقة بين متغيرين نوعيين او أحدهما نوعي والاخر كمي ، او كليهما كميين، ويستخدم هذا الاختبار عندما يكون عدد أزواج القيم n ما بين ($5 \le r_s \le 30$) وتصاغ الفرضية الصفرية كما يلي: $H_0 : \rho_0 = 0$

ومنطقة رفض H_0 تكون كما يلي:

أ) اختبار أحادي من جهة اليمين $r_s > r_{s,\alpha}$ H_0 Reject if

ب)اختبار أحادي من جهة اليسار $r_s < -r_{s,\alpha}$ H_0 Reject if

ت)اختبار من الطرفين $r_s > r_{s,\alpha/2}$ أو $r_s < -r_{s,\alpha/2}$ H_0 Reject if

وتوجد جداول خاصة بهذا الاختبار موجودة في الملحق.

أ-٢) اختبارات أخرى:

اذا كانت n كبيرة (أكبر من حجم جدول الاختبار السابق)

نستخدم اختبار t **بحيث** $t = \frac{r_s}{s}$ مع درجة حرية

$df = n-2$، t توزيع ستيودنت، s الخطأ المعياري ل r_s ،

بحيث $S = (\frac{1-r^2}{n-2})^2$ ، وتوجد صيغة مكافئة لهذه

الاحصاءة (t) باستخدام توزيع F ، حيث $F = \frac{1+|r_s|}{1-|r_s|}$ ، مع

$v_1 = n-2$ و $v_2 = n-2$.

ومع **n كبيرة يمكننا كذلك استخدام التقريب الطبيعي Z**

بحيث $Z = r_s(n-1)^{1/2}$

مثال: البينات التالية تمثل تقديرات لامتحانين (كتابي، شفهي) لثمانية أشخاص تقدموا لشغل منصب ما

ع.إ.كتابي	١٦	١٥	١٤	١٠	١٢	٨	١٣	١٢
ع.إ.شفهي	ممتاز	جيد جدا	جيد	مقبول	جيد	مقبول	جيد	جيد

المطلوب :أ) احسب معامل ارتباط الرتب (r_s)، بين علامات الامتحانين الكتابي و الشفهي؟

ب) اختبر معنوية معامل ارتباط الرتب (r_s)، للفرض القائل انه لا توجد علاقة معنوية بين الامتحانين (استخدم $\alpha = 0.05$)

حل المثال:

أ) حساب معامل ارتباط الرتب (r_s)، نقوم بترتيب كل من الامتحانين في الجدول التالي: (الترتيب اما تصاعديا أو تنازليا للمتغيرين معا)

رتب y	قيم y للامتحان ش	رقم التسلسل		رتب x	قيم x للامتحان ك	رقم التسلسل
١	ممتاز	١		١	١٦	١
٢	جيد جدا	٢		٢	١٥	٢
٤,٥	جيد	٣		٣	١٤	٣
٤,٥	جيد	٤		٤	١٣	٤
٤,٥	جيد	٥		٥,٥	١٢	٥
٤,٥	جيد	٦		٥,٥	١٢	٦
٧,٥	مقبول	٧		٧	١٠	٧
٧,٥	مقبول	٨		٨	٨	٨

لحساب إحصاءة الاختبار (r_s) نقوم باستخدام الجدول التالي:

قيم x للامتحان ك	قيم y للامتحان ش	رتب x R_x	رتب y R_y	$d_i = R_x - R_y$	d_i^2
١٦	ممتاز	١	١	٠	٠
١٥	جيد جدا	٢	٢	٠	٠
١٤	جيد	٣	٤,٥	١,٥−	٢,٢٥
١٠	مقبول	٧	٧,٥	٠,٥−	٠,٢٥
١٢	جيد	٥,٥	٤,٥	١	١
٨	مقبول	٨	٧,٥	٠,٥	٠,٢٥
١٣	جيد	٤	٤,٥	٠,٥−	٠,٢٥
١٢	جيد	٥,٥	٤,٥	١	١
					$\sum_{i=1}^{8} d_i^2 = 5$

إحصاءة الاختبار (r_s)

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6(5)}{8(63)} = 0.94$$

ب)اختبر معنوية معامل ارتباط الرتب (r_s):

ت)صياغة الفرضية تكون كما يلي:

$$H_0 : \rho_0 = 0$$
$$H_1 : \rho_0 \neq 0, \quad \alpha = 0.05$$

من جداول معامل ارتباط لسبيرمان (اختبار من الطرفين)

$r_{s,0/025} = 0.738$, $n = 8$, $\alpha / 2 = 0.025$، منطقة قبول H_0

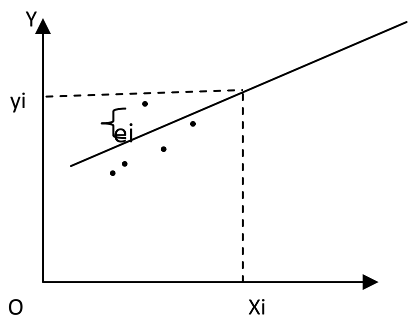
$[-0.738 \quad , \quad 0.738]$

الاستنتاج: بمأن احصاءة الاختبار $r_s = 0.94$ تقع ضمن منطقة رفض فرضية العدم(H_0) أي $(r_s > 0.738)$ وبالتالي نستنتج وجود علاقة معنوية بين المتغيرين أي $(\rho_0 \neq 0)$ وفقا لبيانات العينة عند مستوى $\alpha = 0.05$.

<u>١٠-٣- الانحدار الخطي البسط La Régression linéaire simple:</u>

<u>تمهيد:</u> يعد الانحدار من المواضيع المهمة والأكثر تناولا في ميدان الإحصاء الاستدلالي، باستخداماته الواسعة في شتى الميادين العلمية والاجتماعية والاقتصادية.

<u>تعريفه:</u> هو أداة رياضية تستخدم لتقدير العلاقة بين متغيرين أو أكثر، أحد هذه المتغيرات يسمى متغيرا تابعاvariable dépendant وهو الذي تتأثر قيمته في حالة تغير قيمة المتغير المستقل ويسمى بالمتغير الدال، والآخر يسمى بالمتغير المستقل variable indépendantوهو يؤثر في قيمة المتغير التابع عند تغيره ويسمى بالمتغير المفسر.



لنأخذ الشكل التالي:

لتقدير معالم الانحدار (معامل التقاطع α

والميل β) نستخدم طريقة

المربعات الصغرى (La Méthode de moindre carrées)، إذا

رمزنا بـ $\hat{y}$ للقيمة التي نتنبؤها $y = \mu_{y/x}$ فيمكن كتابة معادلة التنبؤ

على الشكل $\hat{y} = \hat{\alpha} + \hat{\beta}\, x_i$ ، حيث $\hat{\alpha}$ و $\hat{\beta}$ مقدران ل α و β على

الترتيب ، حيث: $\hat{\beta} = \frac{\operatorname{cov}(x, y)}{v(x)}$ و $\hat{\alpha} = \overline{y} - \hat{\beta}$

أو نستخدم طريقة الانحرافات حيث $\hat{\beta} = \frac{S_{xy}}{S_{xx}}$ و $\hat{\alpha} = \overline{y} - \hat{\beta}$

<u>مثال:</u> يمثل الجدول التالي عدد ساعات الدراسة التي أمضاها طالب ما للتحضير للامتحان (x) والعلامات التي حصل عليها:

X	٤	١٠	١٤	٤	٧	١٢	٢٢	١	١٧
Y	٦,٥	١٢	١٣	٧,٥	٩	١٢	١٨	٤,٥	١٦,٥

<u>المطلوب:</u> أوجد معادلة الانحدار الخطي لـy على x.

حل المثال:

$$n=10,\ \sum x_i=100\ ,\ \sum x_i^2=1376\ ,\ \sum y_i=113.5\ ,\ \sum y_i^2=1463.25\ \sum x_i y_i=1385$$

$$S_{xx}=\sum(x_i-\bar{x})^2=\sum x_i^2-\frac{(\sum x_i)^2}{n}=1376-\frac{(100)^2}{10}=376$$

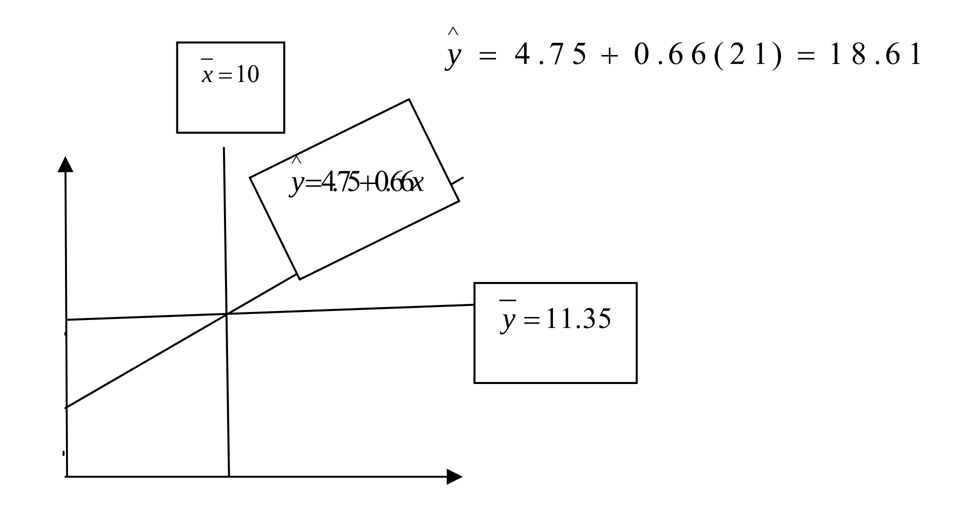
$$S_{xy}=\sum(x_i-\bar{x})(y_i-\bar{y})=\sum x_i y_i-n\bar{x}\bar{y}=1385-\frac{1}{10}(100)(113.5)=250$$

$$\hat{\beta}=\frac{S_{xy}}{S_{xx}}=\frac{250}{376}=0.66,\quad \hat{\alpha}=\bar{y}-\hat{\beta}\bar{x}=113.5-(0.66)(10)=4.75$$

$$\hat{y}=4.75+0.66x$$

يمكن استخدام هذه المعادلة نعرض التنبؤ بقيمة المتغير y من أجل قيمة معينة لـx، فلو فرضنا أن هذا الطالب حضر ٢١ ساعة لهذا الامتحان، فالتنبؤ بالعلامة سيكون كما يلي:

$$\hat{y}=4.75+0.66(21)=18.61$$



١٠-٣-١- اختبار معنوية معامل الانحدار الخطي البسيط:

لاختبار معنوية ميل خط الانحدار (أو معامل الانحدار) ومعامل التقاطع، نضيف افتراض إضافي وهو الخطأ العشوائي (ε) بحيث أنه يتوزع توزيعا طبيعيا بوسط صفر، وتباين (σ^2) نرمز له باختصار $NID(0,\sigma^2)$.

سنطبق فقط اختبار t ، علما انه يوجد كذلك اختبار F لاختبار معنوية الانحدار (تحليل التباين).

اختبار معنوية معامل الانحدار $\hat{\beta}$:

صياغة الفرضية تكون على النحو التالي
$$H_0 : \beta = \beta_0$$
$$H_1 : \beta \neq \beta_0$$

أما عن إحصاءة الاختبار فتكون كما يلي

$$T_0 = \frac{\hat{\beta} - \beta_0}{\sqrt{\hat{\sigma}^2 / S_{xx}}} = \frac{\hat{\beta} - \beta_0}{Se(\hat{\beta})}$$
، تتبع توزيع t مع n-2 درجة حرية

رفض H_0 إذا كان $|t| \geq t_{\alpha/2, n-2}$

وهناك حالة خاصة جدا وهي كثيرة الاستخدام وهي :
$$H_0 : \beta = 0$$
$$H_1 : \beta \neq 0$$

اختبار معنوية معامل التقاطع $\hat{\alpha}$:

صياغة الفرضية تكون على النحو التالي
$$H_0 : \alpha = \alpha_0$$
$$H_1 : \alpha \neq \alpha_0$$

أما عن إحصاءة الاختبار فتكون كما يلي

$$T_0 = \frac{\hat{\alpha} - \alpha_0}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{(\bar{x})^2}{S_{xx}} \right]}} = \frac{\hat{\alpha} - \alpha_0}{Se(\hat{\alpha})}$$

، تتبع توزيع t مع n-2 درجة حرية رفض H_0 إذا كان $|t| \geq t_{\alpha/2, n-2}$

وهناك حالة خاصة جدا وهي كثيرة الاستخدام وهي :

$$H_0 : \alpha = 0$$
$$H_1 : \alpha \neq 0$$

مثال:

نرجع لبيانات المثال رقم() الخاص بعدد ساعات تحضير الطالب للامتحان (x) والعلامات التي حصل عليها (y) ، (نستخدم $\alpha = 0.01$)

المطلوب: اختبار معنوية معلمات نموذج الانحدار $(\hat{\alpha}, \hat{\beta})$

حل المثال:

$$n = 10 \quad , \quad \hat{\beta} = 0.66 \quad , \quad S_{xx} = 376 \quad , \quad \bar{x} = 10$$
$$\sum y_i = 113.5 \quad , \quad \sum x_i y_i = 1385 \quad , \quad \sum y_i^2 = 1463.25$$

أ) **<u>إيجاد الخطأ المعياري للتقدير</u>** $\hat{\sigma}$

يمكن إيجاده وفقا للصيغة التالية :

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - \hat{\alpha} \sum_{i=1}^{n} y_i - \hat{\beta} \sum_{i=1}^{n} x_i y_i}{n-2}} = \sqrt{\frac{1463.25 - (4.75)(113.5) - (0.66)(1385)}{8}} = 1.10$$

ب) **<u>اختبار معنوية معامل الانحدار</u>** $\hat{\beta}$ - :

$$H_0 : \beta = 0 \qquad H_1 : \beta \neq 0 \quad ، \quad \alpha = 0.01$$

$$T_0 = \frac{\hat{\beta}}{\sqrt{\hat{\sigma}^2 / S_{xx}}} = \frac{0.66}{\sqrt{\frac{1.22}{376}}} = 12.19 \text{ ، } t_{0.005,8} = 2.355$$

بمأن t المحسوبة (١٢,١٩) أكبر من القيمة الجدولية (٢,٣٥٥)، وهذا يعني رفض فرض العدم (H_0) مما يدل على معنوية معامل الانحدار($\hat{\beta}$).

ج) **<u>اختبار معنوية معامل التقاطع</u>** $\hat{\alpha}$ - :

$$H_0 : \alpha = 0 \qquad H_1 : \alpha \neq 0 \quad ، \quad \alpha = 0.01$$

$$T_0 = \frac{\hat{\alpha}}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{(\bar{x})^2}{S_{xx}} \right]}} = \frac{4.75}{\sqrt{1.22 \left[\frac{1}{10} + \frac{(10)^2}{376} \right]}} = 7.10 \text{ ،}$$

$$t_{0.005,8} = 2.355$$

بمأن t المحسوبة (٧,١) أكبر من القيمة الجدولية (٢,٣٥٥)، وهذا يعني رفض فرض العدم (H_0) مما يدل على معنوية معامل التقاطع ($\hat{\alpha}$).

تمارين: الارتباط ، الانحدار الخطي البسيط:

تمرين:

إذا كانت لدينا المعلومات التالية:

$$n=15,\ \sum x_i=55\ ,\ \sum x_i^2=220\ ,\ \sum y_i=57.2\ ,\ \sum y_i^2=245\ \ \sum x_i y_i=230.9$$

نفترض أن المتغيرين (x، y) مرتبطين وفقا لنموذج انحدار المطلوب:

أ) وفقا لطريقة المربعات الصغرى، قدر كل من α و β ؟ (معاملي الانحدار).

ب) استخدام المعادلة للتنبؤ إذا كان x= 4.8.

ج) قدر التقدير النقطي للوسط إذا كانت x= 5.4.

د) بافتراض أن قيمة المشاهدة لـ x= 5.4 هي y=6.66.

أحسب القيمة المناظرة للبواقي (ei)؟

هـ) حساب معامل الارتباط لبيرسون؟

١١- الانحدار اللوجستي La Régression Logistique

تمهيد:

إن الهدف من الانحدار الخطي هو تحديد نموذج رياضي يسمح بربط متغير تابع(y)

كمي ومجموعة من المتغيرات المستقلة $x_1 , x_2 ..... x_p$ (متغيرات تفسيرية)، والذي يأخذ الشكل التالي:

$y = b_0 + b_1 x_1 + b_2 x_2 ....$ ، أما نموذج الانحدار اللوجستي يقوم على فرض أساسي هو أن المتغير التابع(y) متغير الاستجابة الذي نهتم بدراسته هو متغير ثنائي يتبع توزيع بيرنولي يأخذ القيمة (١) باحتمال(p) والقيمة (٠) باحتمال q=(1-p) ،أي إلى حدوث الاستجابة وعدم حدوثها.

أمثلة:

١) دراسة تأثير الدخل (x) على صفة ملكية الشخص لعقار معين فإن (y) يأخذ (١) إذا كان الفرد مالكا لهذا العقار و(٠) إذا كان الفرد غير مالك لهذا العقار.

٢) دراسة تأثير عمر الرضيع (x) على اضطراب النوم (y)، فالمتغير التابع

يأخذ (١) إذا كان نومه مضطرب و (٠) إذا كان نومه غير مضطرب.
وهناك من الامثلة الكثيرة في العلوم الاجتماعية.
إذ أن (y) يمثل متغيراً مشاهداً مستمراً وبفرض أن متوسط قيم (y) المشاهدة أو الفعلية عند قيمة معينة للمتغيرx هي $E(y)$ وان المتغير e يمثل الخطأ $e = y - \hat{y}$ فإنه يمكن كتابة النموذج على النحو التالي:

من المعروف في الانحدار أن الطرف الأيمن $E(y/x) = \hat{b}_0 + \hat{b}_1 x$ لهذه النماذج يأخذ قيماً من $(-\infty)$ الى $(+\infty)$ ولكن عندما يكون لدينا متغيران أحدهما ثنائي (y) فإن الانحدار الخطي البسيط لا يكون ملائماً لأن قيمة الطرف الأيمن محصورة ما بين الرقمين (٠،١) وبذلك يكون النموذج غير قابل للتطبيق من وجهة نظر الانحدار العادي، وإن إحدى طرائق حل هذه المشكلة هو إدخال تحويلة رياضية مناسبة على المتغير (y) ، فالنموذج اللوجيستي يعتمد على الدالة التالية :

$$\pi(z) = \frac{1}{1+e^{-z}} = \frac{e^z}{1+e^z} \quad \text{.........................(I)}$$

حيث (e) هو معكوس اللوغاريتم النيبيري ،

Z: تمثل الدالة الخطية المتعددة المتغيرات المستقلة

$$\pi(z) = \frac{1}{1+e^{-(B_0+B_1x_1+B_2x_2+B_3x_3.....)}}$$

هذه الصيغة الأخيرة تحول الى الشكل اللوجيت Logit على النحو التالي:

$$1-\pi(z) = 1-\frac{1}{1+e^{-z}} = \frac{e^{-z}}{1+e^{-z}} = \frac{1}{1+e^z} \quad \text{..........(II)}$$

نقارن بين (I) و (II) فنتحصل على $\frac{\pi(z)}{1-\pi(z)} = e^z$ ، النسبة $\frac{\pi(z)}{1-\pi(z)}$

تسمى نسبة الافضلية او افضلية النجاح (Odds of success)، **للتبسيط نستخدم** $\frac{P}{1-P}$.

إن النسبة $(\frac{p}{1-p})$ أو $(\frac{p}{q})$ عبارة عن مقدار موجب محصور بين $(0,+\infty)$

$0 \le \frac{p}{q} \le +\infty$ ، و بأخذ اللوغاريتم النيبيري للمتغير $\frac{p}{q}$ نتحصل على صيغة تعرف باللوجيت $Logit(P) = Ln(\frac{P}{1-P})$ ، وعليه يمكن كتابة نموذج الانحدار في حاله متغير مستقل واحد: $\ln(\frac{p}{q}) = \hat{b}_0 + \hat{b}_1 x$ ، **وإذا كان لدينا أكثر من متغير مستقل فإن النموذج**

$$\ln(\frac{p}{q}) = \hat{b}_0 + \sum_{i=1}^{k} b_j \hat{x}_{ij}$$ اذا $i = 1,2..........n;;\ j = 1,2..........k$ ،

يسمى هذا النموذج بنموذج الانحدار اللوجستي وتسمى التحويلة $\ln(\frac{p}{q})$ بتحويلة اللوجيت Logit transformation وإن الدالة اللوجستية هي دالة مستمرة تأخذ القيم (٠-١) وتقترب (y) من الصفر كلما اقترب الطرف الأيمن للدالة اللوجستية من $(-\infty)$ وتقترب (y) من الواحد كلما اقترب الطرف الأيسر للدالة اللوجستية من $(+\infty)$، وهي دالة متماثلة عندما يكون الطرف الأيمن لهذه الدالة مساوياً للصفر وتسمى النسبة $\frac{p}{q}$ **نسبة الأفضلية أو أفضلية النجاح** (Odds of success) ، **وان المقدار** $\ln(\frac{p}{q})$ **يسمى لوغاريتم نسبة الأفضلية، او اللوجيت** (Logit) .

ان تقدير معالم نموذج اللوجيت يتم بطريقة (maximum likelihood) **(maximum vraisemblance) ، طريقة الجوازية العظمى ، وهي من أشهر طرق التقدير في** الإحصاء .

١١-١- الاختبارات الاحصائية:

هناك مجموعة من الاختبارات الاحصائية المستخدمة في الانحدار اللوجيستي منها:
اختبار فالد **wald والذي يتبع توزيع كي مربع حيث إحصائية تعطى بالشكل التالي :**

$$wald = \left(\frac{\hat{b}}{se(\hat{b})}\right)^2$$

اختبار Hosmer and Lemeshow يستخدم هذا الاختبار لمعرفة فيما اذا كان النموذج يمثل البيانات بشكل جيد ام لا ، اذ يستخدم اختبار كي مربع χ^2 لحسن المطابقة، لتقييم الفرق بين القيم المشاهدة و المتوقعة ، حيث صياغة الفرضيات كالتالي:

H_0: تساوي الحالات المشاهدة مع الحالات المتنبأ بها ، أي ان النموذج يمثل البيانات بشكل جيد

H_1: عدم تساوي الحالات المشاهدة مع الحالات المتنبأ بها ، أي ان النموذج لا يمثل البيانات بشكل جيد.

مثال:

إذا توفرت لدينا المعلومات التالية حول ١٠٠ رضيع ، تم قياس اضطرابات النوم (y)(مضطرب/ غير مضطرب)، وسن الرضيع (x_1 مقاس بالأشهر)،و(x_2 مشاكل سيكوماتية).

مشاكل صحية x_2	سن الرضيعx_1	اضطراب في النومy	
1	29	1	١
0	14	1	٢
. . .	. . .	. . .	. . .
1	17	1	١٠٠

المطلوب: استخدم الاختبارات التالية:

١) اختبار LRT

٢) اختبار Hosmer and Lemeshow

٣) اختبار فالد **wald.**

حل المثال:

تمت معالجة البيانات بإستخدام برنامج SPSS نسخة ٢٢ وتحصلنا على الجداول التالية:

١١-٢- نتائج تقدير واختبار النموذج اللوجيستى:

الجدول رقم(١)

Iteration History[a,b,c]

Iteration		-2 Log likelihood	Coefficients
			Constant
Step 0	1	118,476	,186
	2	118,476	,187

الجدول رقم(٢)

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients		
			Constant	x1	x2(1)
Step 1	1	111,633	-,487	1,432	,909
	2	111,603	-,510	1,599	,946
	3	111,603	-,511	1,602	,946
	4	111,603	-,511	1,602	,946

الجدول رقم(٣)

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	6,873	2	,032
	Block	6,873	2	,032
	Model	6,873	2	,032

الجدول رقم(٤)

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square

1	111,603[a]	,077	,103

الجدول رقم(٥)

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	9,715	7	,205

الجدول رقم(٦)

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	x1	1,602	1,336	1,437	1	,231	4,961
	x2(1)	,946	,454	4,337	1	,037	2,576
	Constant	-,511	,351	2,112	1	,146	,600

التفسير:

من خلال الجدول رقم (١) كانت قيمة $-2LL_{(1)}=118.476$ الخاصة بالثابت بعد تكرارين (deux itérations)، أما الجدول رقم (٢) كانت قيمة $-2LL_{(2)}=111.603$ الخاصة بالثابت وx1 ،x2 و بعد اربع تكرارات .

كي مربع للنموذج يطلق عليه ايضا Log likelhood (LR) ، حيث يحسب كما يلي:

LR= 118.476-111.603= 6.03 ، $\chi^2 = 2[\log_e L_0 - \log_e L_1]$
الجدول رقم (٣)

وهذه القيمة معنوية لان ٠،٠٣٢ < ٠،٠٥ عند درجة حرية (٢) (متغيرين تفسيريين) ، ونستنتج على الاقل متغير من x1,x2 يؤثر على المتغير التابع (y1).

تقييم جودة التوفيق للنموذج:

في نموذج الانحدار اللوجيستي يستعاض عن معامل التحديد (R^2) الذي يستخدم لمعرفة ملاءمة نماذج الانحدار المقترحة لبيانات الدراسة بإحصائيتي التوفيق (Nagelkerke R Square)

و (Cox & Snell R Square) اللتين لهما هدف الاحصاءة (R^2) في الانحدار الخطي **المتعدد** الجدول رقم(٤) وتحسب كما يلي:

$$R^2 = 1 - [\frac{L_0}{L_1}]^{\frac{2}{n}}$$

$$\tilde{R}^2 = \frac{R^2}{R^2_z}$$

$$R^2_z = 1 - (L_0)^{(\frac{2}{n})}$$

L_0 : دالة الجوازية العظمى في حالة النموذج يحوي على الثابت فقط

L_1 : دالة الجوازية العظمى في حالة النموذج يضم جميع المتغيرات التوضيحية

n : حجم العينة.

يمثل الجدول رقم (٤) جودة التوفيق للنموذج ، نجد ان قيمة (-2LL) للنموذج الحالي هي (١١١,٦٣٣) هي اقل من (-2LL) الخاصة بالنموذج الذي يحوي الثابت فقط البالغة(118.476) مما يدل على جودة النموذج الذي يحتوي كل المتغيرات التوضيحية عن الذي يحتوي على الثابت فقط ، أما قيمة احصاءة Nagelkerke R Square و Cox & Snell R Square فهما يهدفان الى تحديد نسبة التباين المفسرة في نموذج الانحدار اللوجيستي ، وبهذا لهما نفس احصاءة (R^2 معامل التحديد) في الانحدار المتعدد، وبالنظر الى $\tilde{R}^2$ (الجدول رقم (٤)) نجد انها تساوي على التوالي (٠،١٠٣)، (٠,٠٧٧) تقريبا من التباين في متغير التابع ثم نفسره بالمتغيرات التوضيحية.

بالنسبة للجدول رقم (٥) اختبار Hosmer and Lemeshow Test كانت (sig=0.205) وهي اكبر من مستوى المعنوية المحدد (α=0.05) قبول H_0 مما يشير الى ان الحالات المشاهدة تتساوى مع الحالات المتنبأ بها، أي ان النموذج يمثل البيانات بشكل جيد.

أما باقي المعلومات التي تخص معالم المتغيرات التوضيحية المقدرة واحصاءة wald والتي تحسب وفق العلاقة التالية $wald = \left(\frac{\hat{b}_i}{se(\hat{b}_i)}\right)^2$ ،

وقيمة مستوى المعنوية الخاص بتلك الاحصاءة ، ولوغاريتم نسبة المفاضلة فتم تلخيصها في الجدول رقم (٦) ،نلاحظ ان العمود (B) يحتوي على معاملات النموذج المرفق وهي بوحدات Log odds وتكون معادلة النموذج على النحو التالي:

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -0.511 + 1.602\,x_1 + 0.946\,x_2$$

للعودة الى الجدول رقم (٦) نلاحظ ان قيمة X_2 (مشاكل صحية ٠،٩٤٦) وتبين ان احصاءة wald البالغة ٤،٣٣٧ ومن خلال (sig=0 .037<0.05)، معنوية جيدة للمعلمة المقدرة ، وهذا يشير الى اهمية عامل المشاكل الصحية وتأثيره على اضطراب النوم ، اما بالنسبة لنسبة الافضلية (ExpB) للمعلمة (٢،٥٧٦) فتعني إذا تغير المتغير التوضيحي من مشاكل صحية (بالنسبة للرضع) الى غير الذين يعانون من مشاكل صحية، فأن نسبة التغير في افضلية (اضطراب في النوم) تزداد بمقدار (٢،٥٧٦) .

وبالرغم من ان قيمة المعلمة X_1 صغيرة (١،٦٠٢) الا أن القيمة المناظرة لاحصاءة wald

غير معنوية (sig=0 .231>0.05) اي ان هذا المتغير ليس له تأثير معنوي في النموذج.

<u>قائمة المراجع المعتمدة:</u>

I) <u>الكتب باللغة العربية:</u>

١- شيبات أحمد - الإحصاء الوصفي، ترجمة حسان زواش، دار الهدى- عين مليلة ،٢٠٠١.

٢- جلاطو جيلالي ، الإحصاء التطبيقي، دار الخلدونية، ط١، ٢٠٠٧/١٤٢٨.

٣- ليبسشتر سيمور ، الاحتمالات، ترجمة سامع داود، الدار الدولية للنشر، القاهرة، ٢٠٠٤.

٤- بن يخلف مصطفى ، الاحتمالات والإحصاء الرياضي، دار النشر المغربية، ط١، المغرب، ١٩٧٥.

٥- ماندري جان بول ، الاحتمالات، ترجمة خالد سعد الله، opu، الجزائر، ١٩٩٩.

٦- عبد الكريم البشير ، إحصاء٢، دار الكتاب العربي، الجزائر، ٢٠٠٦.

٧- طعمة ياسين ، حسين حنوش إيمان ، أساليب الإحصاء التطبيقي، ط١، دار صفاء للنشر والتوزيع، عمان، ٢٠٠٩/١٤٣٠.

٨- عمر قاسم عزات ، مبادئ الاحتمالات والإحصاء، منشورات جامعة دمشق، ١٩٩٥.

٩- مصطفى عبد الحفيظ - نظرية الاحتمالات، ج١، ط٢، opu،
الجزائر، ٢٠٠٤,

١٠-القاضي دلال، عبد الله سهيلة، البياتي محمود ، الاحصاء للإداريين
والاقتصاديين، ط١، دار الحامد، عمان، ٢٠٠٤.

II) الكتب باللغة الفرنسية:

1١- Pierre Boulay Jean, Staistique mathématique
ellipses, paris 2010.

1٢- Redjal- K, Cours de probabilités, opu, Alger,
2004.

1٣- Spiegel Muerray , théorie et Applications de la
statistique, traduction Alain ergas et jean français
Marcotorchino, 17 tirage Mc Craw-hill Paris- 1991.

1٤- Saporta Gilbert, probabilité, Analyse des données
et statistique, 2 édition, technique , Paris 2006.

١٥- Charles pupion Pierre, statistique pour la gestion, 3éme édition, Duno, Paris, 2012.

١٦- stafford jean, Bodson paul, L'analyse multivariée avec spss,presses de l'université du québec,2006.

١٧- Tribout Brigitte, statistique pour économistes et gestionnaires, pearson Education, Paris, 2007.

18- Borsali fethi, statistique médicales et biologiques, Ellipses, Paris, 2010.

III) الكتب باللغة الانجليزية:

19- Miller Irwin, Jhon Preund, probability md statistics for engineers, 2nd édition, prentice-Hall, New jersey.

20- C.R. Héathcote, probability, elements of the mathematical Theory, First published, By page Bros (Nor wish) great Britain, 1971.

21- Z.Raqab Mohamed , Awad adnan, Azzam Mufid, principles of statistics, second edition, Academic for publishing and Distributing, Amman, 2005.

٢2- New bold Paul, Statistics for Bussiness and economics, fourthedition, prentice-Hall International.Inc, New jersey, 1995.

٢3- Roussas George, Introduction to probability and statistical Inference, Academic press, Boston, 2003.

24- Montgomery Douglas, George Runger, Applied Statistics and probability for Engineers, third edition, Jhon Wiley and Sons, Inc, New york T_1, T_2, 2002.

25- Dowdy Shirley, Stanley Weardon, Daniel. Chilko, Statistics for research, thirde dition, Jhon Wiley and Sons,Inc, publication, New jersey, 2004.

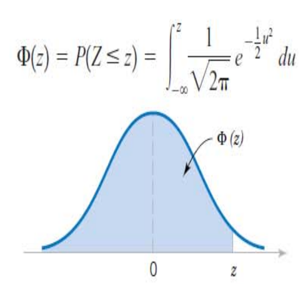
الدوريات

٢٦- قاسم عبد الرزاق، أثر بعض المتغيرات في الاصابة بمرض اللثة بإستخدام نموذج الانحدار اللوجسيتي ، مجلة العلوم الاقتصادية، جامعة البصرة، العدد السابع كانون الاول ٢٠١١.

٢٧-غانم عدنان، الجاعوني فريد خليل، استخدام تقنية الانحدار اللوجسيتي ثنائي الاستجابة في دراسة أهم المحددات الاقتصادية والاجتماعية لكفاية دخل الاسرة ، دراسة تطبيقية على عينة عشوائية من الأسر في محافظة دمشق، مجلة جامعة دمشق للعلوم الاقتصادية والقانونية، مجلد ٢٧-العدد الاول ٢٠١١.

جداول إحصائية

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du$$

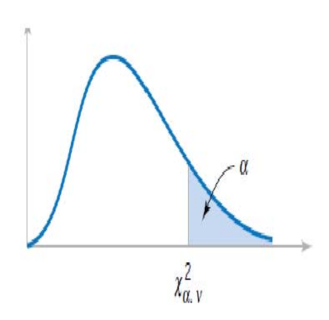


Cumulative Standard Normal Distribution (*continued*)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267

0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289

3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

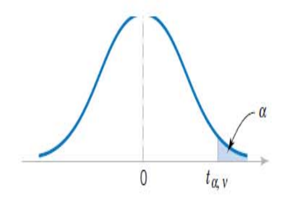


Percentage Points $\chi^2_{\alpha,\nu}$ of the Chi-Squared Distribution

ν \ α	.995	.990	.975	.950	.900	.500	.100	.050	.025	.010	.005
1	.00+	.00+	.00+	.00+	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55

ν											
7	.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

ν = degrees of freedom.



Percentage Points $t_{\alpha,v}$ of the t-Distribution

v \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437

12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

TABLE A.11 TABLE DU COEFFICIENT DE CORRÉLATION DES RANGS DE SPEARMAN ENTRE DE DEUX VARIABLES INDÉPENDANTES

Valeurs r de R_s ayant une probabilité α d'être dépassée en valeur absolue

$P(|R_s| > r) = \alpha$

n \ α	0.50	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
4	0.600	1.000	1.000						
5	0.500	0.800	0.900	1.000	1.000				
6	0.371	0.657	0.829	0.886	0.943	1.000	1.000		
7	0.321	0.571	0.714	0.786	0.893	0.929	0.964	1.000	1.000
8	0.310	0.524	0.643	0.738	0.833	0.881	0.905	0.952	0.976
9	0.267	0.483	0.600	0.700	0.783	0.833	0.867	0.917	0.933
10	0.248	0.455	0.564	0.648	0.745	0.794	0.830	0.879	0.903
11	0.236	0.427	0.536	0.618	0.709	0.755	0.800	0.845	0.873
12	0.224	0.406	0.503	0.587	0.671	0.727	0.776	0.825	0.860
13	0.209	0.385	0.484	0.560	0.648	0.703	0.747	0.802	0.835
14	0.200	0.367	0.464	0.538	0.622	0.675	0.723	0.776	0.811
15	0.189	0.354	0.443	0.521	0.604	0.654	0.700	0.754	0.786
16	0.182	0.341	0.429	0.503	0.582	0.635	0.679	0.732	0.765
17	0.176	0.328	0.414	0.485	0.566	0.615	0.662	0.713	0.748
18	0.170	0.317	0.401	0.472	0.550	0.600	0.643	0.695	0.728
19	0.165	0.309	0.391	0.460	0.535	0.584	0.628	0.677	0.712
20	0.161	0.299	0.380	0.447	0.520	0.570	0.612	0.662	0.696
21	0.156	0.292	0.370	0.435	0.508	0.556	0.599	0.648	0.681
22	0.152	0.284	0.361	0.425	0.496	0.544	0.586	0.634	0.667
23	0.148	0.278	0.353	0.415	0.486	0.532	0.573	0.622	0.654
24	0.144	0.271	0.344	0.406	0.476	0.521	0.562	0.610	0.642

25	0.142	0.265	0.337	0.398	0.466	0.511	0.551	0.598	0.630
26	0.138	0.259	0.331	0.390	0.457	0.501	0.541	0.587	0.619
27	0.136	0.255	0.324	0.382	0.448	0.491	0.531	0.577	0.608
28	0.133	0.250	0.317	0.375	0.440	0.483	0.522	0.567	0.598
29	0.130	0.245	0.312	0.368	0.433	0.475	0.513	0.558	0.589
30	0.128	0.240	0.306	0.362	0.425	0.467	0.504	0.549	0.580
31	0.126	0.236	0.301	0.356	0.418	0.459	0.496	0.541	0.571
32	0.124	0.232	0.296	0.350	0.412	0.452	0.489	0.533	0.563
33	0.121	0.229	0.291	0.345	0.405	0.446	0.482	0.525	0.554
34	0.120	0.225	0.287	0.340	0.399	0.439	0.475	0.517	0.547
35	0.118	0.222	0.283	0.335	0.394	0.433	0.468	0.510	0.539
36	0.116	0.219	0.279	0.330	0.388	0.427	0.462	0.504	0.533
37	0.114	0.216	0.275	0.325	0.383	0.421	0.456	0.497	0.526
38	0.113	0.212	0.271	0.321	0.378	0.415	0.450	0.491	0.519
39	0.111	0.210	0.267	0.317	0.373	0.410	0.444	0.485	0.513
40	0.110	0.207	0.264	0.313	0.368	0.405	0.439	0.479	0.507

(*suite et fin*) TABLE DU COEFFICIENT DE CORRÉLATION DES RANGS
DE SPEARMAN DE DEUX VARIABLES INDÉPENDANTES
Valeurs r de R_s ayant une probabilité α d'être dépassée en valeur absolue
$P(|R_s| > r) = \alpha$

n \ α	0.50	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
41	0.108	0.204	0.261	0.309	0.364	0.400	0.433	0.473	0.501
42	0.107	0.202	0.257	0.305	0.359	0.395	0.428	0.468	0.495
43	0.105	0.199	0.254	0.301	0.355	0.391	0.423	0.463	0.490
44	0.104	0.197	0.251	0.298	0.351	0.386	0.419	0.458	0.484
45	0.103	0.194	0.248	0.294	0.347	0.382	0.414	0.453	0.479
46	0.102	0.192	0.246	0.291	0.343	0.378	0.410	0.448	0.474
47	0.101	0.190	0.243	0.288	0.340	0.374	0.405	0.443	0.469
48	0.100	0.188	0.240	0.285	0.336	0.370	0.401	0.439	0.465
49	0.098	0.186	0.238	0.282	0.333	0.366	0.397	0.434	0.460
50	0.097	0.184	0.235	0.279	0.329	0.363	0.393	0.430	0.456
52	0.095	0.180	0.231	0.274	0.323	0.356	0.386	0.422	0.447
54	0.094	0.177	0.226	0.268	0.317	0.349	0.379	0.414	0.439
56	0.092	0.174	0.222	0.264	0.311	0.343	0.372	0.407	0.432
58	0.090	0.171	0.218	0.259	0.306	0.337	0.366	0.400	0.424
60	0.089	0.168	0.214	0.255	0.300	0.331	0.360	0.394	0.418
62	0.087	0.165	0.211	0.250	0.296	0.326	0.354	0.388	0.411
64	0.086	0.162	0.207	0.246	0.291	0.321	0.348	0.382	0.405
66	0.084	0.160	0.204	0.243	0.287	0.316	0.343	0.376	0.399
68	0.083	0.157	0.201	0.239	0.282	0.311	0.338	0.370	0.393
70	0.082	0.155	0.198	0.235	0.278	0.307	0.333	0.365	0.388

72	0.081	0.153	0.195	0.232	0.274	0.303	0.329	0.360	0.382
74	0.080	0.151	0.193	0.229	0.271	0.299	0.324	0.355	0.377
76	0.078	0.149	0.190	0.226	0.267	0.295	0.320	0.351	0.372
78	0.077	0.147	0.188	0.223	0.264	0.291	0.316	0.346	0.368
80	0.076	0.145	0.185	0.220	0.260	0.287	0.312	0.342	0.363
82	0.075	0.143	0.183	0.217	0.257	0.284	0.308	0.338	0.359
84	0.074	0.141	0.181	0.215	0.254	0.280	0.305	0.334	0.355
86	0.074	0.139	0.179	0.212	0.251	0.277	0.301	0.330	0.351
88	0.073	0.138	0.176	0.210	0.248	0.274	0.298	0.327	0.347
90	0.072	0.136	0.174	0.207	0.245	0.271	0.294	0.323	0.343
92	0.071	0.135	0.173	0.205	0.243	0.268	0.291	0.319	0.339
94	0.070	0.133	0.171	0.203	0.240	0.265	0.288	0.316	0.336
96	0.070	0.132	0.169	0.201	0.238	0.262	0.285	0.313	0.332
98	0.069	0.130	0.167	0.199	0.235	0.260	0.282	0.310	0.329
100	0.068	0.129	0.165	0.197	0.233	0.257	0.279	0.307	0.326

Pour $n > 100$ on admet que R_s est distribué comme $LG\left(0\,;\frac{1}{\sqrt{n-1}}\right)$.